



## House price estimation using an eigenvector spatial filtering approach

McCord, M., McCord, J., Davis, P., Haran, M., & Bidanset, P. (2019). House price estimation using an eigenvector spatial filtering approach. *International Journal of Housing Markets and Analysis*, 13(5), 845-867. <https://doi.org/10.1108/IJHMA-09-2019-0097>

[Link to publication record in Ulster University Research Portal](#)

### Published in:

International Journal of Housing Markets and Analysis

### Publication Status:

Published online: 14/11/2019

### DOI:

[10.1108/IJHMA-09-2019-0097](https://doi.org/10.1108/IJHMA-09-2019-0097)

### Document Version

Author Accepted version

### General rights

Copyright for the publications made accessible via Ulster University's Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [pure-support@ulster.ac.uk](mailto:pure-support@ulster.ac.uk).



## House price estimation using an eigenvector spatial filtering approach

Journal:	<i>International Journal of Housing Markets and Analysis</i>
Manuscript ID	IJHMA-09-2019-0097.R1
Manuscript Type:	Research Paper
Keywords:	Eigenvector spatial filtering, Spatially varying coefficients, Hedonic price analysis, Spatial dependency, Spatial Autocorrelation, House Prices

SCHOLARONE™  
Manuscripts

Emerald allows authors to deposit their AAM under the Creative Commons Attribution Noncommercial International Licence 4.0 (CC BY-NC 4.0). To do this, the deposit must clearly state that the AAM is deposited under this licence and that any reuse is allowed in accordance with the terms outlined by the licence. To reuse the AAM for commercial purposes, permission should be sought by contacting [permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)

## House price estimation using an eigenvector spatial filtering approach

### Abstract

*Purpose:* Numerous geo-statistical methods have been developed to analyse the spatial dimension and composition of house prices. Despite these advances, spatial filtering remains an under-researched approach within house price studies. This paper examines the spatial distribution of house prices using an eigenvector spatial filtering (ESF) procedure, to analyse the local variation and spatial heterogeneity.

*Methods:* Using 2,664 sale transactions over the one year period Q3 2017 to Q3 2018, an eigenvector spatial filtering approach is applied to evaluate spatial patterns within the Belfast housing market. This method consists of using geographical coordinates to specify eigenvectors across geographic distance to determine a set of spatial filters. These convey spatial structures representative of differing spatial scales and units. The filters are incorporated as predictors into regression analyses to alleviate spatial autocorrelation. This approach is intuitive, given that detection of autocorrelation in specific filters and within the regression residuals can be markers for exclusion or inclusion criteria.

*Results:* The findings show both robust and effective estimator consistency and limited spatial dependency - culminating in accurately specified hedonic pricing models. The findings show that the spatial component alone explains 14.6% of the variation in property value, whereas 77.6% of the variation could be attributed to an interaction between the structural characteristics and the local market geography expressed by the filters. This methodological approach reduced short-scale spatial dependency and residual autocorrelation resulting in increased model stability and reduced misspecification error.

*Originality:* Eigenvector-based spatial filtering is a less known but suitable statistical protocol that can be used to analyse house price patterns taking into account spatial autocorrelation at varying (different) spatial scales. This approach arguably provides a more insightful analysis of house prices by removing spatial autocorrelation, both objectively and subjectively, to produce reliable, yet understandable regression models which do not suffer from traditional challenges of serial dependence or spatial mis-specification. This approach offers property researchers and policy makers an intuitive but comprehensible approach for producing accurate price estimation models which can be readily interpreted.

**Keywords:** Eigenvector spatial filtering, Spatially varying coefficients, Hedonic price analysis, house prices, spatial dependency, spatial autocorrelation

### Introduction

There has been increasing emphasis placed on the accuracy of house price estimation and its role in informing urban policy from healthy communities to liveable spaces and connected places. It also retains significance for a wide range of economic activity. In light of this importance, the accuracy, stability and defenceability of house price models is crucial for wider property market analysis and the robust insights into local housing market conditions (Bourassa, Cantoni, and Hoesli, 2010; Seya et al., 2014). Recent advances in geographic information systems (GIS) and geo-statistical and spatial econometric approaches, both parametric and non-parametric, have advanced house price analysis, as they consider the geographically distributed properties of spatial data and identify spatial dependence and

spatial heterogeneity. This has, by-and-large, been a consequence of the awareness of the potential error-bias contained within traditional hedonic models when not acknowledging the spatial heterogeneity of pricing effects (Lu, *et al.*, 2014; Helbich and Griffith, 2015). This increased awareness has propagated considerable interest in accounting for spatial non-stationarity and dependence within hedonic analysis (McCord *et al.*, 2014), challenging the assumption of a constant price function uniformly across a housing market area – an assumption that does not conform to the operation of and complexity of housing market mechanics congruent with urban economic theory (McMillen and Redfearn, 2010).

The advancement of geo-statistical approaches, and the discipline of spatial econometrics, has introduced numerous discussions regarding the confounding effects of space (LeSage and Pace 2009), and primarily the role of spatial autocorrelation – one of the important characteristics of spatial data (Anselin, 1988). This causes the inflation of Type I errors in the significance tests of correlation and regression analyses which can overestimate the degrees of freedom, reduce confidence intervals and result in errors in statistical inference under a null hypothesis, bias coefficients and lead to inappropriate conclusions (LeSage and Pace, 2009; Páez, Fei, and Farber, 2008). Pertinently, many spatial analysis techniques thereby employ model-based statistical inference, the dependability of which is based upon the correctness of assumptions about a model's error term, namely its randomness and independence of observations.

Despite the appealing methodological and practical advantages of these more localised spatial approaches, there remains disagreement within real estate applications as to which method is superlative (Griffith, 2008; Helbich and Griffith, 2015). Numerous approaches have been developed and advocated over the past two decades which have incorporated an eclectic range of spatial effects in order to account for locational effects in hedonic price modelling. These approaches, such as Simultaneous Autoregressive Regression, Spatial Expansion methods, Spatial Lag Models (which incorporate spatial structural instability) and spatial drift models, make use of the spatial characteristics of variables to improve results through reduced error terms and spatial independence (Gao *et al.*, 2006). One prevailing method within property analysis is the Geographically weighted regression (GWR) approach introduced by Fotheringham *et al.*, (1998; 2002), which has become a major approach for explicitly accounting for spatial heterogeneity, using spatially varying coefficients. Nonetheless, this method has not been without its criticisms. Páez *et al.* (2011) and Wheeler and Tiefelsdorf (2005) have indicated that the basic GWR model typically suffers from multicollinearity issues, and also assumes the same degree of spatial smoothness for each coefficient using bandwidth criteria which can affect the model results (Bidanset and Lombard, 2014; Bidanset *et al.*, 2017). There has been refinements to the approach, such as the incorporation of Principal Component Analysis (PCGWR), as a methodology to help remove multicollinearity from inclusionary neighbourhood and locational determinants. More recently there have been augmented approaches suggested incorporating weighting matrices for spatial, temporal and property characteristics, such as the GTCWR approach of Bidanset *et al.* (2018).

One approach which remains relatively unknown, yet emerging as an alternative procedure to address spatial dependence, is the Spatial filtering approach - a method developed in order to obtain enhanced and robust results in a spatial data analysis framework by removing spatial dependency (Griffiths, 1996, 2003; Tiefelsdorf and Griffith, 2007). In its basic format, the

eigenvector spatial filtering (ESF) method is an approach that captures spatial dependence applying map pattern variables obtained from spatial connectivity information, using the Moran (1950) coefficient. This is achieved through the decomposition of a spatial variable/characteristic into trend, a spatially structured stochastic signal, and random noise. In essence, it separates spatially structured random components from both trend and random noise, culminating in leads to sounder statistical inference and useful visualization (Griffith, 1996; 2008; Griffith and Chun, 2014; Helbich and Griffith, 2015). This separation procedure involves eigenfunctions of the matrix version of the numerator of the Moran Coefficient (Griffith and Chun, 2014). Therefore, the application of an ESF approach is to create a spatially structured random component, as captured by a linear combination of selected eigenvectors, in order to mitigate potential error bias through limiting autocorrelation within the residuals.

In this regard, the ESF methodology has been increasingly utilised in a variety of regression settings. The basic model is identical to the standard ordinary (OLS) and generalized (GLS) least squares linear regression models, and therefore, it is easily implemented (Griffith, 2003; Murakami and Griffiths, 2015). As discussed by Tiefelsdorf and Griffith (2007), spatial filtering addresses this from a semi-parametric perspective by generating synthetic explanatory variables reflecting the data's spatial structure. This increases flexibility into model processing in order to analytically decompose a variable into underlying (spatial) components which provides a synthetic variate (the spatial filter) to visualise any spatial autocorrelation contained within a geo-referenced variable (Murakami and Griffiths, 2015). According to Thayn and Simanis (2013) and Franzese and Hays (2014) this approach produces unbiased parameter estimates, reduces spatial misspecification error; increases model fit; increases the normality of model residuals and can increase the homoscedasticity of model residuals spatial dependence and spatial spill-over effect.

This study develops an ESF model (as per Griffith 2003) for the Belfast housing market and thereby models geographically varying relationships using a subset of eigenvectors extracted from a spatial weights matrix as synthetic control variables in a regression model specification. This aims to remove spatial dependence and increase standardised regression models estimation reliability. This approach is seen as furnishing a parsimonious solution to the geographically varying linear regression coefficients problem, a better understanding of multicollinearity, and improved accounting for spatial autocorrelation. More pertinently, it provides professionals with a readily understandable methodology for applying spatial analysis in a more standardised and explainable hedonic framework.

**Literature**

Eigenvector Spatial Filtering, as outlined by Murakami and Griffith (2015), has become increasingly popular for the understanding of spatial phenomena, with applications increasing in light of its practicality, readily adaptable process and integration within classical regression based techniques which are considered transparent and understandable. The approach has been adopted across a number of scientific disciplines for understanding spatial interaction, ecological and economic processes and more specifically land-use and housing market analysis. A core stand of this literature has tended to examine the comparative



performance of spatial filtering approaches with other locally weighted regression or spatial expansion methodologies and its effectiveness for analysing spatial autocorrelation within regression models.

An early study conducted by Griffith and Peres-Neto (2006), from an ecology perspective, analysed two differing spatial filtering approaches, to help investigate and explain the geographic variability associated with ecological communities. Their results demonstrate the usefulness of eigenfunctions in spatial modelling - specifically that the manifestation of spatial predictors can be easily incorporated into conventional regression models for analysis. Indeed, the authors showed that an important advantage of the spatial filtering methodology over other spatial approaches is that they provide a flexible tool that allows the full range of general and generalized linear modelling theory to be applied to ecological and geographical problems in the presence of nonzero spatial autocorrelation. Analogous findings are evident in the work of Blanchet *et al.* (2008) who, also in an ecological context, investigated the distribution of species, and explicitly the direction of an asymmetric process controlling species distributions along a biogeographical gradient or network, using an eigenfunction-based spatial filtering technique. Comparing the ESF with traditional Moran's eigenvector maps (MEM) analysis within a simulation framework they find that the ESF is superior for producing unbiased coefficient estimations.

With regards to spatial interaction analysis, Chun (2008) tested the assumption of independence among interaction flows engaged in spatial interaction modelling, in the context of U.S. interstate migration flows for measuring network autocorrelation. Undertaking a Stepwise incorporation of eigenvectors, which are extracted from a network link matrix to capture the network autocorrelation in a Poisson regression, the results showed that estimated regression parameters in the spatial filtering interaction model become more intuitively interpretable. Similarly, Fischer and Griffith (2008) compared two approaches, the spatial interaction gravity model and the eigenfunction-based spatial filtering approach, to deal with the issue of spatial autocorrelation amongst flow residuals across 112 European regions. In line with Chun (2008) their findings showed the ESF to be more intuitive. This was also evident in the study of Chun and Griffith (2011) who employed the eigenvector spatial filtering technique to analyse network autocorrelation among migration flows structured through multiple time spans. The findings showed improved model fitting and more intuitive parameter estimates.

In a wider economic context, Crespo-Cuaresma and Feldkircher (2013) also employed spatial filtering to measure the spatial uncertainty of income convergence in Europe using a dataset of income per capita growth and 50 potential determinants for 255 NUTS-2 European regions. The authors reveal that spatial linkages (matrices) comprise an important effect on the estimates of the parameters attached to the model covariates and that income convergence in Europe is influenced by spatially correlated growth spill-overs. Similarly, Patuelli *et al.* (2011) examined regional performance related to unemployment rates in 439 NUTS-3 German districts. They employed a spatial filtering model to unemployment rates in Germany using the derived spatial filters as explanatory variables in a panel modelling framework. Their results show that the computed spatial filters account for most of the residual spatial autocorrelation in the data. In a follow-up study, Patuelli *et al.* (2012) investigate the dynamic adjustment process of unemployment to the study of regional unemployment persistence, in order to account for spatial heterogeneity and/or spatial autocorrelation in both the levels and the dynamics of unemployment. They also employ the use of spatial filtering as a substitute for fixed effects within a panel estimation framework in order to incorporate region-specific information that generates spatial autocorrelation, frees up degrees of freedom and

simultaneously corrects for time-stable spatial autocorrelation in the residuals. The authors find widely heterogeneous, but generally high, persistence in regional unemployment rates, signifying that ESF helps provide insights about the spatial patterns in regional adjustment processes.

Griffith and Chun (2014) investigated regional population forecasting for South Korea by incorporating spatial autocorrelation in a generalized linear mixed model framework coupled with eigenvector spatial filtering to capture spatial autocorrelation, namely the complex map pattern portraying spatial dependence that is latent in population counts, and preserves it in regional forecasts of population. The authors find that empirical evaluations of the short run population forecasts indicate that using an ESF to describe spatially structured random effects coupled with a spatially unstructured random effects term furnishes good annual county-level geographic resolution predictions. A further study inspecting regional inequality in China's Guangdong region by Liao and Wei (2015) applied a spatial filtering method in order to eliminate spatial dependence and quantify the extent to which spatial effects have contributed to regional inequality at multiple scales. The results indicated the effect of strengthening spatial dependence with the authors concluding that spatial filtering as a tool helps improve the understanding of complex spatial phenomena.

The approach has also been utilised within the confines of criminal analysis, where Chun (2014) using eigenvector spatial filtering analysed the space-time crime incidents relating to vehicle burglary in Texas, USA, between 2004-2009 within a Poisson generalized linear mixed model specification using ESF. The author shows the approach to be an efficient tool for furnishing robust estimates. In terms of crime mapping and spatial crime analysis, Helbich and Arsanjani (2015) employ ESF as a method for undertaking spatio-temporal mapping to uncover time-invariant crime patterns. Their results suggest that local and regional geography significantly contributes to the explanation of crime patterns. Furthermore, they show annual space-time eigenvectors to indicate spatio-temporal patterns persisting over time. Their findings show that spatial filtering successfully absorbs latent autocorrelation and, therefore, prevents parameter estimation bias whilst increasing the explanatory power of the regression analysis.

Moniruzzaman and Paez (2012) apply spatial filtering for examining urban design analysis and the implications of accessibility to transit for the city of Hamilton, Canada. Employing a logistic regression approach which they highlight are sensitive to overdispersion and spatial error autocorrelation which can result in misleading inference and erroneous policy prescriptions, they show that using spatial filters improved the model inference and accounted for over-dispersion and spatial autocorrelation. In a similar study, Wang, Kockelman and Wang (2013) explored the application of spatial filtering for regression model estimation for transportation land use and land value estimation. Using case studies and appraised values for private properties the authors analysed the effectiveness of spatial filtering in comparison to spatial autoregressive (SAR) models. Their findings showed the SF approach offers increased goodness of fit statistics and more reliable marginal effects of policy variables and other covariates, in comparison to more conventional SAR-based models. Murakami *et al.* (2017) also compare their eigenvector spatially varying coefficient model to GWR to examine land prices for flood hazards in Japan. Using a Monte Carlo simulation technique their study reveals outperformance of geographically weighted regression (GWR) models in terms of the accuracy of parameter estimates and computational time. Further, the authors highlight that the developed model has spatially varying

coefficients which have a different degree of spatial smoothness which is a challenge for conventional GWR.

The application of ESF in housing market analysis has been relatively limited, despite that fact that housing policy requires the recognition of spatial heterogeneity in housing prices to account for local settings. McCord *et al.* (2013) examined a number of spatially based modelling frameworks encompassing more traditional approaches (OLS) to more complex spatial filtering methods to estimate rental values within the Belfast housing market, UK. Their findings revealed that GWR showed increased accuracy, albeit nominal, in predicting marginal price estimates relative to eigenvector filtering and other spatial techniques. Nonetheless, they noted that the high level of segmentation across localised pockets of the housing market needed further analytical insights as the smooth bandwidth did not adequately capture this whereas spatial filtering did, concluding that soft segmentation modelling approaches are essential for understanding rental gradients. More recently, Helbich and Griffith (2016) examined the application of the ESF model in relation to the spatial variation of house prices in a comparative assessment between locally weighted methodologies. The findings showed the ESF to depict a more localized pattern of the parameter estimates without local smoothing. Moreover they revealed the ESF to be less affected by multicollinearity issues between the local parameter estimates than the other approaches. The authors do however show that whilst ESF demonstrates superiority for in-sample explanatory power and prediction accuracy, the weighted regression approaches exhibit slightly better out-of-sample estimations. Nonetheless, they conclude by advocating for the consideration of ESF as a valuable alternative for real estate research that allows going beyond normal probability models.

Overall, the foregoing analysis suggests that spatial filtering comprises relative advantages for understanding complex spatial dependence and autocorrelation across a wealth of ecological and regional economic problems and more latterly housing market analysis. As illustrated by Thayn and Simanis (2013), OLS models whilst comprising well-known limitations for spatial analysis, are useful and easily interpreted, and the assumptions, strengths, and weaknesses of these models are well studied and understood. Accordingly, they advocate that spatial filtering is a powerful geographic method that should be applied to regression based models that use geographic data. In this regard, and in light of a modest number of extant studies examining house prices spatially employing this technique, this paper uses the spatial filtering technique to analyse house price patterns across the Belfast housing market.

## Data and Methodology

The analysis is conducted using 2,664 sales transactions over the 12 month period (Q3 2017 to Q3 2018) after undergoing a data mining and cleansing exercise to remove outliers. The data was integrated into a GIS platform to append property address information in order to derive absolute location coordinates ( $X$ ,  $Y$ ) required for the spatial modelling exercises<sup>1</sup>. The independent variables are based on the structural characteristics of the properties, including the era of construction, property typology, property size and the number of bedrooms,

---

<sup>1</sup> The data was exported into SAM, an integrated computational platform tool for spatial analyses (See: Rangel *et al.*, 2010). Processing time for generation of the ESFs equated to approximately 17 minutes for the study sample size. This time accounts for the truncation distance calculation of the maximum connectivity between all sampling units under the minimum spanning tree criterion and the filter selection extraction response determination.



reception rooms and whether the property comprises a garage (Table I). Where applicable, the categorical variables are transformed into their binary state. This process is undertaken to indicate the absence or presence of a categorical effect that may be expected to shift the outcome (Kleinbaum *et al.*, 1988).

<<< **Insert Table I Variable Descriptions**>>>

A summary of the descriptive statistics for the data is presented in Table II. The sample mean property price is £171,781 which reveals a high dispersion and positive skew (Figure I). As a consequence, the logarithmic of sale price was calculated in order to standardise the house price variable and satisfy the statistical assumptions of normality for modelling purposes. The average floor size equates to 106m<sup>2</sup>, again displaying a high variance.

<<< **Insert Figure I Frequency distributions of sale price and the logarithmic of sale price**>>>

<<< **Insert Table II Descriptive Statistics**>>>

**Methodology**

Initially, topology-based eigenvector based spatial filtering rests upon the seminal work of de Jong, Sprenger, and van Veen (1984), who pioneered studying and applying the relationship between eigenvalues and the Moran's I coefficient to avoid spatial autocorrelation and regression misspecification identified by earlier authors (Cliff and Ord, 1973). In this regard, the ESF method, as developed by Griffith (2000), utilises geographical coordinates which are subject to an eigen analyses of geographical distances to establish a set of spatial filters (eigenvectors) expressing the spatial structure of the region at different spatial scales. In other words, spatial filtering addresses heterogeneity in behaviours through interacting eigenvalues and systematic covariates (Wang *et al.*, 2011). This process exploits eigenvector decomposition techniques, thereby extracting orthogonal and uncorrelated numerical components from a given contiguity matrix (Patuelli *et al.*, 2012), which also emerges in the numerator of the Moran Coefficient statistic. The Moran ESF is based on the Moran coefficient which is a spatial dependence diagnostic statistic formulated as follows:

$$MC = \frac{N \mathbf{y}' \mathbf{M} \mathbf{C} \mathbf{M} \mathbf{y}}{\mathbf{1}' \mathbf{C} \mathbf{1} \mathbf{y}' \mathbf{M} \mathbf{y}} \tag{1}$$

where **1** is an  $N \times 1$  vector of ones, **y** is an  $N \times 1$  vector of variable values, **C** is an  $N \times N$  connectivity matrix whose diagonal elements are zero, and  $\mathbf{M} = \mathbf{I} \mathbf{N} - \mathbf{1} \mathbf{1}' / N$  is an  $N \times N$  matrix for double centring, where **IN** is an  $N \times N$  identity matrix. Notably, **M** is replaced with  $\mathbf{M} \mathbf{X} = \mathbf{I} \mathbf{N} - \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'$  if **y** is a residual vector of a linear regression model. As highlighted by Griffith (2003) and Murakami et al. (2017), the MC is positive if the sample values in **y** display positive spatial dependence, and negative if they display negative spatial dependence. The *l*-th eigenvector of **MCM**, **e<sub>l</sub>**, describes the *l*-th map pattern explained by MC, while the set of eigenvectors of **MCM**, **E<sub>full</sub>** = {**e<sub>1</sub>**, ..., **e<sub>N</sub>**}, provides all the possible distinct map

pattern descriptions of latent spatial dependence, with each magnitude being indexed by its corresponding eigenvalue (Griffith, 2003).

As illustrated above, the resulting eigenvectors become mutually uncorrelated and orthogonal, with each mimicking a certain degree of latent spatial autocorrelation (SAC), representing global to local patterns (Tiefelsdorf and Griffith, 2007). Accordingly, the eigenvector corresponding to the first eigenvalue,  $\mathbf{e}_1$ , is the set of real values that has the largest positive MC (depicting the maximum positive spatial dependence) achievable by any set of real numbers for the spatial arrangement defined by  $\mathbf{C}$ . The second eigenvector,  $\mathbf{e}_2$ , is the set of real values that has the largest positive MC that is uncorrelated with and orthogonal to  $\mathbf{e}_1$ , and  $\mathbf{e}_N$  is the set of numerical values that has the largest negative MC (depicting the maximum negative spatial dependence) achievable that is uncorrelated with and orthogonal to  $\mathbf{e}_1, \dots, \mathbf{e}_{N-1}$  (Griffith 2003). The set of eigenvectors of  $\mathbf{MCM}$ ,  $\mathbf{E}_{full} = \{\mathbf{e}_1, \dots, \mathbf{e}_N\}$ , furnishes all possible distinct map pattern descriptions of latent spatial dependence, with each level being indexed by an MC that is proportional to its corresponding eigenvalue. The basic linear model of ESF is:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \quad (2)$$

where  $\mathbf{y}$  is a  $N \times 1$  vector of response variable values;  $\mathbf{E}$  is a  $N \times L$  matrix composed of a subset of  $L$  eigenvectors ( $L < N$ ) from  $\mathbf{E}_{full}$ ;  $\boldsymbol{\varepsilon}$  is a  $N \times 1$  vector of disturbances;  $\boldsymbol{\beta}$  and  $\boldsymbol{\gamma} = [y_1, \dots, y_L, \dots, y_L]$  are parameter vectors whose sizes are  $K \times 1$  and  $L \times 1$ , respectively;  $\sigma^2$  is a variance parameter; and  $\mathbf{0}$  is a  $N \times 1$  vector of zeros. Equation (2) thus is a semiparametric model, with  $\mathbf{M} = \mathbf{I} - \mathbf{1}\mathbf{1}'/N$ , [Eq. (2)] an approximation of a standard spatial lag model, whereas, when  $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ , [Eq. (2)] an approximation of the spatial error model (Tiefelsdorf and Griffith 2007). When  $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ , the eigenvectors are mutually uncorrelated as well as uncorrelated with  $\mathbf{X}$ .

As outlined by Griffith (2003), the  $L$  eigenvectors in  $\mathbf{E}$  are selected by: (1) Eigenvectors representing inconsequential levels of spatial dependence are removed, and (2) significant eigenvectors are chosen using a stepwise selection method. Commonly, step (1) is conducted by removing eigenvectors whose eigenvalues are small or of the wrong nature using the adjusted  $R^2$  as the objective function, with step (2) achieved by maximizing model accuracy or minimizing residual spatial dependence using the MC. Notably, Eq. (2) is identical to the standard linear regression model, thus step (2) can be conducted by using ordinary least squares (OLS) estimation-based stepwise methods. OLS estimators of  $\boldsymbol{\beta}$  and  $\boldsymbol{\gamma}$  are given as:

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{E} \\ \mathbf{E}'\mathbf{X} & \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{E}'\mathbf{y} \end{bmatrix} \quad (3)$$

In addition, Griffith (2008:2761) further augments the basic linear model by introducing interaction terms between the selected eigenvectors and the predictors to model spatially varying coefficients as opposed to using the final EVs to correct for SAC on a global level. Accordingly, the extension takes the following form:

$$\hat{Y} \approx (\beta_0 \mathbf{1} + \sum_{K=1}^{K_0} E_{k0} \beta_{k0}) + \sum_{P=1}^P (\beta_p \mathbf{1} + \sum_{K_p=1}^{K_p} E_{K_p} \beta_{K_p}) \cdot X_p + \boldsymbol{\varepsilon}$$

(4)

where  $\hat{Y}$  is the  $n \times 1$  vector of prices,  $X_p$  is a  $n \times 1$  vector of independent variable  $p$  ( $p=1,2,3, \dots, P$ ),  $E_k$  is the  $K_p$  EV ( $k=1,2,3, \dots, K$ ) that describes the variable  $p$ ,  $\beta_0$ ,  $\beta_{k0}$ ,  $\beta_{kp}$  are estimated regression coefficients, and  $\varepsilon$  is an independent and identically distributed error term. Note that the element-wise matrix multiplication and the interaction terms are given by  $\beta_{kp} \cdot X_p$ .<sup>2</sup> The first part of the equation represents the spatially varying intercept, and the second part represents the spatially varying coefficients. After rearranging, the regression coefficients constitute the global impact, while the individual EVs mimic local modifiers of these global effects across space:

$$Y = \beta_0 1 + \sum_{p=1}^P X_p \cdot 1 \beta_p + \sum_{k=1}^K E_k \beta_{E_k} + \sum_{p=1}^P \sum_{k=1}^K X_p \cdot E_k \beta_{pE_k} + \varepsilon \quad (5)$$

In practice, the outlined procedure is challenging due to a large set of covariates and interaction terms, eventually larger than the available number of degrees of freedom. Griffith (2008) originally proposed forward variable selection to find significant interactions, but this procedure is computationally slow (Seya *et al.*, 2014). In order to identify the most relevant interactions in a parsimonious manner, the Akaike information criterion (AIC) is used for evaluation purposes, which considers the model fit and penalizes less parsimonious models. Finally, in order to obtain the final and mappable coefficients, all ESF model parts with common attributes are collected and then factored out in order to determine its spatially varying coefficient (Griffith, 2008).

### Partial Regression approach

The semi-partial regression is used to express the specific portion of variance explained by a given independent variable within the regression analysis (Abdi, 2007). Indeed, this approach is primarily employed for non-orthogonal linear regression to assess the specific effect of each independent variable on the dependent variable (Larsen and McCleary, 1972), where the partial regression coefficient or partial slope coefficient value is dependent upon the other independent variables included in the regression equation. Within the traditional OLS setting, the multiple regression is extended to find a set of partial regression coefficients  $\beta_k$  such that the dependent variable could be approximated as well as possible by a linear combination of the independent variables. Therefore, a predicted value, denoted  $\hat{Y}$ , of the dependent variable is obtained as:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_k X_k \dots \beta_K X_K \quad (6)$$

The value of the partial coefficients are found using ordinary least squares (OLS). It is often convenient to express the multiple linear regression equation using matrix notation. In this framework, the predicted values of the dependent variable are collected in a vector denoted  $\hat{y}$  and are obtained using:

<sup>2</sup> The parameters are estimated by means of OLS.

$$by = Xb \text{ with } b = (X^T X)^{-1} X^T y$$

(7)

### Model development and Eigenvector Filter identification

As highlighted in the methodology, the ESF approach performs more efficiently in a parsimonious model - as with all spatial modelling architectures. In this regard, there has been continued debate within the field of spatial econometrics relating to the standard statistical testing approach when applied to non-experimental, usually broad-scale spatial data, especially with respect to model selection procedures (Cohen, 1994). The debate centres around the inclusion of additional parameters for increasing model predictability and the obvious increase in (geo-statistical) model complexity. Consequently, we apply an initial multi-model inference procedure to reduce model complexity, remove potential variance inflation and multicollinearity<sup>3</sup> without compromising model explanation, predictability and stability. This is achieved through the assessment of the minimisation of the Akaike Information Criteria (AICc)<sup>4</sup>. This approach deemed the most parsimonious model, based on 3,095 OLS models tested, to include 12 parameters, excluding the Garage variable and the number of bedrooms thereby providing an  $R^2$  of 63% and Adjusted  $R^2$  of 61.3% (Table III).

<<<Insert Table III OLS Model Selection procedure sorted by Akaike Information Criterion>>>

#### *Eigenvector Filter identification*

As discerned previously, the purpose of ESF is to remove spatial trends in the response variable, in this instance house prices. The eigenvectors were calculated using geographic coordinates applying a truncated distance function allowing the maximum distance connectivity to connect all sampling units under a minimum tree criterion. The filters were pre-selected based on minimisation of the residual Moran's I ( $p=0.05$ ) across the distance class boundaries (lower and upper) guided by the logarithmic of house price as the response variable. When employed as regressors, the eigenvectors function as proxies for missing explanatory variables, through a parsimonious set of 'candidate' eigenvectors<sup>5</sup>. The residuals obtained constitute the spatially filtered component of the geo-referenced variable examined, computed on the basis of a modified spatial weights matrix. In total, 758 eigenvector filters were established to 'filter' for spatial autocorrelation.

The Eigenvector spatial filtering technique can adopt both a pre-selection criterion and a judicious selection of eigenvectors, as the number of filters appointed tends to increase with both level of linear regression residual spatial autocorrelation and the number of areal units. The spatial filters are subsequently examined with the extraction of the filters to be utilised in the regression modelling undertaken using a filter selection criteria with minimisation of the

<sup>3</sup> This procedure estimates the relative quality of the models for the given set of data, relative to each of the other models premised on the relative information lost by a given model: the less information a model loses, the higher the quality of that model. This therefore estimates the trade-off between the 'goodness of fit' of the model and the simplicity of the model.

<sup>4</sup> The AIC(c) statistic is based on the maximum likelihood of estimating parameters,  $\beta_i$ , where the probability of the observed data would be as large as possible (Burnham and Anderson, 2002), computed as its small sample corrected version as this is asymptotic to the standard version: See De Smith et al. (2007) for a full discussion

<sup>5</sup> selected from the  $n$  eigenvectors, on the basis of their  $MI$  values.

residuals is achieved based on a local Moran’s  $I$  statistic. This automated step for filter pre-selection was further scrutinised to test for potential model customisation regarding the trade-off between increasing the model explanation ( $R^2$ ), the AICs and any potential increases in residual variance inflation. This minimises the residual short-distance spatial autocorrelation and reduces the level of residual autocorrelation. In addition, this step thereby ensured model optimality and model stability whilst further encompassing the assessment of each spatial filters spatial correlogram and the variance of the log-price estimation as demonstrated in Figure II.

**<<<Insert Figure II Spatial filter diagnostic testing>>>**

In total, 66 spatial filters were extracted and retained for the regression modelling. As evidenced in Figure III, which displays a sample of extracted filters<sup>6</sup>, each filter extracted presents a detailed representation of the spatial patterns which can have a different degree of spatial structure, smoothness and geographically varying relationship with house prices. Notably the spatial structure becomes more ‘localised’ when displaying the filters with smaller eigenvalues culminating in more localized parameter surfaces given the reduced truncation distances.

**<<<Insert Figure III Spatial Representation of Filters extracted>>>**

**OLS and Partial regression results**

Having identified the optimal spatial filters using the AIC selection criterion and pragmatic investigative scrutiny of each filter, the spatial filters are used as independent predictors in multiple and partial regression analyses to mitigate spatial autocorrelation and error bias. In total, three regression models are specified to account for location namely, the inclusion of spatial filters (*Model I*) and secondly, the linear combination of the filters suppressed into one location coefficient (*Model II*) and an interaction spatial filter model (*Model III*). Overall, all models display good explanation and performance with an Adjusted  $R^2$  of 0.769, 0.760 and 0.776 with the  $F$ -tests confirming model validity. As observed in Table IV, the distance classifications with their respective accompanying residual Moran’s  $I$  value shows that this has reduced to a low level, with only the immediate short-distance (Dis. Class 1) showing the presence of any noteworthy small-scale spatial autocorrelation within the residuals.

**<<<Insert Table IV Moran’s  $I$  residual autocorrelation within the spatial matrix>>>**

Tables V and VI summarise the estimated coefficients across the models. The coefficients for all models infer that for every  $m^2$  increase in property size this equates to a 0.5% and 0.6% increase in price respectively<sup>7</sup>. With regards to property type, both terrace and apartments exhibit statistically significant negative coefficients with the detached coefficient revealing a 23.1% and 25.98% increase ( $p<.01$ ) in models I and II. The property age coefficients reveal negative coefficients across all age categories with the exception of new build properties in both models which exhibits a 14.7% and 11.4% percentage effect. In terms of spatial effects, Model I presents the eigenvector spatial filters which reveal geographically varying regression coefficients, both positive and negative, representing regional patterns and

<sup>6</sup> the histogram of spatial filter selection residuals evident in Appendix I

<sup>7</sup> Measured using the  $\epsilon(\beta)$ -1 transformation as discussed by Halvorsen and Palmquist (1980).



aggregation effects. Model II further presents an overall linear combination of the extracted filters showing the more global coefficient to be significant ( $p < .001$ ).

In terms of partial effects, the OLS model is further constructed to define predictor sets, to examine overlap in explanation under identified predictor set categories. The property characteristics (floor area; type; age) are separated from the spatial filter explanatory parameters to derive a series of additive models which partition the explanation into unique and shared components. For model I, the property characteristics (Predictor set *A*) explain 61.5% of the variation in house prices, with the pure spatial dimension represented by the spatial filters (Predictor set *B*) explaining 36.2% with an overall total explanation of 76.7%. In terms of unique contribution, 39.1% of the variation in property price is explained solely by the physical characteristics, with 15.2% exclusively by the spatial characteristics and 21.1% shared explanation between the property characteristics and the spatial component (Table V). For model II, there is a marginally reduced level of explanation for the linear combination of the spatial filters (35.3%) with the unique explanation of the spatial dimension decreasing to 13.9%. However the shared explanation between the predictor sets *A* and *B* marginally increased by 0.4% to 21.4% signalling that the linear combination of the spatial filters nominally increases the level of multicollinearity between the physical and spatial attributes. Overall, the results show that the filter selection inclusion has effectively eliminated residual spatial autocorrelation, whilst not overlapping in any considerable manner with the physical characteristics within the partial regression analysis.

#### <<< Insert Table V Regression (Partial) models>>>

Further refinement of the model specification through the inclusion of interacting property type with age variables, as illustrated in Table VI, slightly increases the model explanatory power, albeit marginally. The findings display an eclectic and varied pricing effect across the property type and age interactions - symbolic of housing market structure, segmentation and heterogeneity, explaining 63.4% of the variation in house prices. In terms of the spatial component, the predictor set B encompassing the spatial filters shows an explanation of 40.3% culminating in an overall  $R^2$  of 78.1%. Further insights as to each specific contribution to the model, 37.7% of the variation in property price is explained solely by the physical characteristics, with 14.6% by the extracted filters with the shared explanation equating to 25.8%.

#### <<<Insert Table VI Interactive (Partial) Regression Model>>>

In terms of spatial representation, Figure IV reveals the estimation and residuals for each respective model. The results present some localised price patterns characteristic of the topographical nature of the housing market structure and two distinctive areas of market segmentation. To the north-west of the Belfast housing market, the results show a lower pricing structure with a number of enclaves towards the centre of this area forming the lowest house prices in the overall market. Towards the south of the market, there is evidence of small pockets of elevated pricing clusters or hot spots. Finally, the model residuals exhibit few instances of elevation and relative stability with limited residual spatial autocorrelation evident across the housing market geography.

#### <<<Insert Figure IV Regression model estimates and residuals>>>

## Conclusion

Spatial analysis within house price studies has evolved significantly over the past two decades with numerous methodologies having emerged to examine the spatial patterns, heterogeneity and dependency of house prices. This has been fuelled by the ever-growing interest, and indeed importance of housing market policy within urban analyses and policymaking. Despite these advances, eigenvector spatial filtering remains a largely unknown spatial approach within house price estimation studies, despite its increasing application across a variety of other disciplines. The ESF approach provides a foundation for including location within the confines of a traditional regression approach to produce stable, reliable estimates devoid of spatial dependency. This is achieved through the generation of synthetic explanatory variables reflecting the data's spatial structure. The findings emanating from this study show the effectiveness of applying this methodology to house price sales data for the Belfast housing market, revealing that the spatial filters can be observed as linear combinations of the eigenvectors, and can be regarded as patterns of independent spatial dimensions, culminating in the almost complete elimination of residual spatial autocorrelation and therefore mitigating parameter estimation bias.

The findings exhibit localised parameter surfaces capable of mapping local parameter estimates which does not assume the constant bandwidth or nearest neighbours assumption necessary for other techniques such as GWR, whilst not facing multicollinearity issues between the local coefficients. This provides market professionals and policymakers with a more readily and understandable methodology for applying spatial analysis in a more standardised and explainable hedonic framework. Moreover, the findings, using a partial regression approach in order to isolate the spatial effects, revealed that the unique explanation of the spatial dimension accounts for 14.6% of price variation across the Belfast housing market with limited overlap with the physical characteristics.

The approach has the capacity to aid practitioners in the property taxation field by allowing a decomposition between more intangible spatial elements contributing to value and more tangible physical characteristics. In terms of overall accuracy both are included and defensible. The spatial similarities and differences can be assessed and mapped – depicting similar and dissimilar areas. This can potentially be illustrated by comparable transactions from these areas. The remainder can be illustrated by a more simple model with more understandable parameter estimates for features such as presence of garage or floor area. These estimates will not vary spatially, but across relatively small distances, the cost of providing them also does not vary greatly. This facilitates discussion of the relative contribution of tangible and intangible value factors in a way which may be beneficial for discussions with non-technical consumers of model derived appraisals – such as tax payers and assessment tribunal members. This functionality may be of use as a result in terms of ‘explainability’ of model derived value estimates, which can be crucial in terms of defending assessments and underpinning the operation of effective and efficient property taxes for raising vital public finance.

There is one cautionary note, however. Whilst the ESF approach does offer property researchers and policy makers an intuitive but comprehensible approach for producing accurate price estimation models which can be readily interpreted, one explanation for the ESF approach remaining somewhat of an outcast within house price studies relates to its lack of a ‘user friendly’ interface. As it is premised on eigenvector extraction from a neighbourhood connectivity matrix, this necessitates a large set of interaction terms, as

evidenced in this research, which produced 758 eigenvector spatial filters as local parameter estimates. Dealing with this can be both computationally and time intense. It is not, however, outside the bounds of complexity or opacity of comparable spatial and machine learning approaches such as GWR or ANN and perhaps deserves to be included in discussions regarding advanced alternatives to traditional regression analysis for understanding housing markets and for applications seeking to harness such understanding, such as automated valuation modelling for mortgage lending, or mass appraisal of residential values for property taxation purposes.

## References

- Abdi, H. (2007). Part (semi partial) and partial regression coefficients. In: Salkind, N (Ed.) (2007) *Encyclopedia of measurement and statistics*, 736-740, Thousand Oaks (CA): Sage.
- Bidanset, P. E., and Lombard, J. R. (2014). The effect of kernel and bandwidth specification in geographically weighted regression models on the accuracy and uniformity of mass real estate appraisal. *Journal of Property Tax Assessment and Administration*, 11(3), 5-14.
- Bidanset, P. E., Lombard, J. R., Davis, P., McCord, M., and McCluskey, W. J. (2017). Further Evaluating the Impact of Kernel and Bandwidth Specifications of Geographically Weighted Regression on the Equity and Uniformity of Mass Appraisal Models. In *Advances in Automated Valuation Modeling* (pp. 191-199). Springer, Cham.
- Bidanset, P., McCord, M., Lombard, J. A., Davis, P., and McCluskey, W. (2018). Accounting for Locational, Temporal, and Physical Similarity of Residential Sales in Mass Appraisal Modeling: Introducing the Development and Application of Geographically, Temporally, and Characteristically Weighted Regression (GTCWR). *Journal of Property Tax Assessment and*, 14(2), 5-13.
- Blanchet, F. G., Legendre, P., and Borcard, D. (2008). Modelling directional spatial processes in ecological data. *ecological modelling*, 215(4), 325-336.
- Bourassa, S., Cantoni, E., and Hoesli, M. (2010). Predicting house prices with spatial dependence: a comparison of alternative methods. *Journal of Real Estate Research*, 32(2), 139-159.
- Chun, Y. (2008). Modeling network autocorrelation within migration flows by eigenvector spatial filtering. *Journal of Geographical Systems*, 10(4), 317-344.
- Chun, Y., and Griffith, D. A. (2011). Modeling network autocorrelation in space-time migration flow data: an eigenvector spatial filtering approach. *Annals of the Association of American Geographers*, 101(3), 523-536.
- Chun, Y. (2014). Analyzing space-time crime incidents using eigenvector spatial filtering: an application to vehicle burglary. *Geographical Analysis*, 46(2), 165-184.
- Crespo Cuaresma, J., and Feldkircher, M. (2013). Spatial filtering, model uncertainty and the speed of income convergence in Europe. *Journal of Applied Econometrics*, 28(4), 720-741.
- Chun, Y., and Griffith, D. A. (2014). A quality assessment of eigenvector spatial filtering based parameter estimates for the normal probability model. *Spatial Statistics*, 10, 1-11.
- De Jong, P., Sprenger, C., and Van Veen, F. (1984). On extreme values of Moran's I and Geary's c. *Geographical Analysis*, 16(1), 17-24.
- Fischer, M. M., and Griffith, D. A. (2008). Modeling spatial autocorrelation in spatial interaction data: an application to patent citation data in the European Union. *Journal of Regional Science*, 48(5), 969-989.
- Gao, X., Asami, Y., and Chung, C. J. F. (2006). An empirical evaluation of spatial regression models. *Computers and Geosciences*, 32(8), 1040-1051.
- Griffith, D. A. (2000). Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses. *Linear Algebra and its Applications*, 321(1-3), 95-112.
- Griffith, D. A. (1996). Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo-referenced data. *Canadian Geographer/Le Géographe canadien*, 40(4), 351-367.

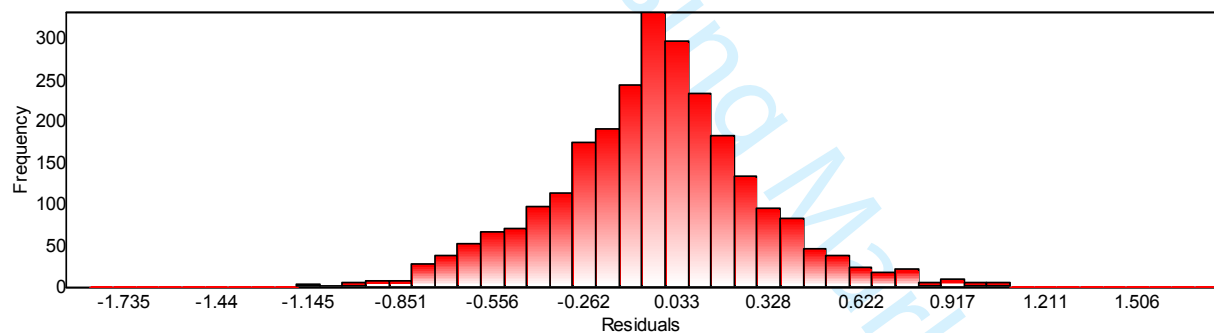
- Griffith DA (2003) Spatial autocorrelation and spatial filtering: Gaining understanding through theory and scientific visualization. Springer, Berlin.
- Griffith, D. A. (2008). Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment and Planning A*, 40(11), 2751-2769.
- Griffith DA (2009) Modeling spatial autocorrelation in spatial interaction data: empirical evidence from 2002 Germany journey-to-work flows. *J Geogr Syst* 11(2):117–140
- Griffith DA (2010) Spatial filtering. In: Fischer MM, Getis A (eds) *Handbook of applied spatial analysis*. Springer, Berlin, pp 301–318
- Griffith DA (2011) Visualizing analytical spatial autocorrelation components latent in spatial interaction data: an eigenvector spatial filter approach. *Comput Environ Urban Syst* 35(2):140–149
- Griffith DA, Chun Y (2013) Spatial autocorrelation and spatial filtering. In: Fischer MM, Nijkamp P (eds) *Handbook of regional science*. Springer, Berlin, pp 1477–1507
- Griffith, D. A. (2009). Spatial autocorrelation. *International encyclopedia of human geography*, 2009, 308-316.
- Griffith, D., and Chun, Y. (2014). Spatial autocorrelation and spatial filtering. *Handbook of regional science*, 1477-1507.
- Griffith, D. A., and Chun, Y. (2014). An eigenvector spatial filtering contribution to short range regional population forecasting. *Economics and Business Letters*, 3(4), 208-217.
- Halvorsen, R. and Palmquist, R. (1980), The interpretation of dummy variables in semi-logarithmic equations, *American Economic Review*, Vol. 70 No. 3, pp. 474-475.
- Helbich, M., and Jokar Arsanjani, J. (2015). Spatial eigenvector filtering for spatiotemporal crime mapping and spatial crime analysis. *Cartography and Geographic Information Science*, 42(2), 134-148.
- Helbich, M., and Griffith, D. A. (2016). Spatially varying coefficient models in real estate: Eigenvector spatial filtering and alternative approaches. *Computers, Environment and Urban Systems*, 57, 1-11.
- Jackson, L. E., Kose, M. A., Otrok, C., and Owyang, M. T. (2016). Specification and estimation of bayesian dynamic factor models: A monte Carlo analysis with an application to global house price comovement. In *Dynamic Factor Models* (pp. 361-400). Emerald Group Publishing Limited.
- Larsen, W. A., & McCleary, S. J. (1972). The use of partial residual plots in regression analysis. *Technometrics*, 14(3), 781-790.
- LeSage J.P. and Pace R. K. (2009) *Introduction to spatial econometrics*. Chapman and Hall/CRC, Boca Raton
- LeSage J.P. and Pace R. K. (2014) The biggest myth in spatial econometrics. *Econom* 2(4):217–249.
- Liao, F. H., and Wei, Y. D. (2015). Space, scale, and regional inequality in provincial China: A spatial filtering approach. *Applied Geography*, 61, 94-104.
- Maza, A., and Villaverde, J. (2008). A note on the need to account for spatial dependence: A case of migratory flows in Spain. *The review of regional studies*, 38(1), 105.
- McCord, M., Davis, P. T., Haran, M., McIlhatton, D., and McCord, J. (2014). Understanding rental prices in the UK: a comparative application of spatial modelling approaches. *International Journal of Housing Markets and Analysis*, 7(1), 98-128.
- Murakami, D., Yoshida, T., Seya, H., Griffith, D. A., and Yamagata, Y. (2017). A Moran coefficient-based mixed effects approach to investigate spatially varying relationships. *Spatial Statistics*, 19, 68-89.
- Moniruzzaman, M., and Páez, A. (2012). Accessibility to transit, by transit, and mode share: application of a logistic model with spatial filters. *Journal of Transport Geography*, 24, 198-205.
- Páez, A., Farber, S., and Wheeler, D. (2011). A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. *Environment and Planning A*, 43(12), 2992-3010.
- Páez, A., Long, F., and Farber, S. (2008). Moving window approaches for hedonic price estimation: an empirical comparison of modelling techniques. *Urban Studies*, 45(8), 1565-1581.



- Patuelli, R., Schanne, N., Griffith, D. A., and Nijkamp, P. (2012). Persistence of regional unemployment: Application of a spatial filtering approach to local labor markets in Germany. *Journal of Regional Science*, 52(2), 300-323.
- Patuelli, R., Griffith, D. A., Tiefelsdorf, M., and Nijkamp, P. (2011). Spatial filtering and eigenvector stability: space-time models for German unemployment data. *International Regional Science Review*, 34(2), 253-280.
- Rangel, T.F.; Diniz-Filho, J.A.F. and Bini, L.M. (2010), SAM: A comprehensive application for Spatial Analysis in Macroecology. *Ecography*, 33: 1-5.
- Seya H, Tsutsumi M, Yamagata Y (2014a) Weighted-average least squares applied to spatial econometric models: a Monte Carlo investigation. *Geographical Analysis* 46(2):126–147
- Seya, H., Tsutsumi, M., Mendez, S. A., and Vega, A. M. (2012). Application of model averaging techniques to spatial hedonic land price models. *Econometrics: New Research*, 63-88.
- Seya, H., Murakami, D., Tsutsumi, M., and Yamagata, Y. (2015). Application of LASSO to the Eigenvector Selection Problem in Eigenvector-based Spatial Filtering. *Geographical Analysis*, 47(3), 284-299.
- Thayn, J. B., and Simanis, J. M. (2013). Accounting for spatial autocorrelation in linear regression models using spatial filtering with eigenvectors. *Annals of the Association of American Geographers*, 103(1), 47-66.
- Tiefelsdorf, M., and Griffith, D. A. (2007). Semiparametric filtering of spatial autocorrelation: the eigenvector approach. *Environment and Planning A*, 39(5), 1193-1221.
- Wang, Y., Kockelman, K. M., and Wang, X. C. (2013). Understanding spatial filtering for analysis of land use-transport data. *Journal of Transport Geography*, 31, 123-131.
- Wheeler, D., and Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems*, 7(2), 161-187.

## Appendix I

*Histogram of spatial filter selection residuals*





Tables and Figures

<<< Table I Variable Descriptions>>>

Variable	Description	Type
Sale Price	Transaction price	C
In(PPrice)	Log of transaction price	C
Floor area	Size of Floor area in m <sup>2</sup>	C
Property Type	Type of property (transformed to binary e.g. 1 if Terr; 0 otherwise)	B
Property Age	Age of property (transformed to binary e.g. 1 if Pre1919; 0 otherwise)	B
Bedrooms	Number of bedrooms (transformed to binary e.g. 1 if 1 bed; 0 otherwise)	B
Reception	Number of reception rooms (transformed to binary e.g. 1 if 1 reception; 0 otherwise)	B
Garage	Garage present (transformed to binary e.g. 1 if Garage; 0 otherwise)	B
Sale period	Date of sale period (transformed to binary e.g. 1 if Q3 2017; 0 otherwise)	B

NB. C = continuous; B = binary.

<<<Table II Descriptive Statistics>>>

Variable	Sale Price	Log price	Size	Beds	Reception
N	2,664	2,664	2,664	2,664	2,664
Minimum	25,000	10.127	27	1	1
Maximum	1,600,000	13.911	528	7	3
1st Quartile	95,000	11.462	78	3	1
Median	140,000	11.849	98	3	1
3rd Quartile	205,000	12.231	118	4	2
Mean	171,780	11.871	105.735	3.072	1.481
Std. Deviation	120,066	0.587	44.773	0.887	0.683
Skewness	2.48	0.324	2.533	0.604	1.281
Kurtosis	109.32	-0.307	135.225	14.619	15.877

<<<Table III OLS Model Selection procedure sorted by Akaike Information Criterion>>>

Model Num.	Variables (#)	No. Vars	R <sup>2</sup>	Cond. Num.	AICc	Delta AICc	L(gi x)	AICc wi
Mod #308	1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	12	0.63	2.429	2094.528	0	1	0.044
Mod #149	1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14	13	0.63	3.474	2095.547	1.019	0.601	0.026
Mod #177	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14	14	0.63	3.307	2096.307	1.779	0.411	0.018
R <sup>2</sup>	0.63							
Adj. R <sup>2</sup>	0.613							

\*Parameter estimates averaged across 3,095 OLS models using Akaike Weights (AICc wi).  
#1:FloorArea;#2:Reception;#3:Beds;#4:TER;#5:SEMI;#6:DET;#7:APT;#8:Pre1919;#9:Interwar;#10:Postwar;#11:Earlymodern;#12:Postl  
980;#13 New Build;#14No Garage.

<<<Table IV Moran's I residual autocorrelation within the spatial matrix>>>

Dis. Class	Count	Distance	Model I Res. M i	Model II Res. M i	Model III Res. M i
1	159302	338.338	0.148	0.17	0.146
2	352576	1015.013	0.038	0.062	0.041
3	453968	1691.688	-0.037	-0.018	-0.037
4	501812	2368.363	-0.062	-0.049	-0.063
5	507420	3045.038	-0.047	-0.042	-0.038
6	530700	3721.713	0.028	0.013	0.028
7	605894	4398.388	0.044	0.034	0.045
8	663664	5075.063	0.011	0.01	0.01
9	701008	5751.738	-0.005	-0.011	-0.004
10	668360	6428.413	-0.009	-0.024	-0.009

NB. the first 10 of 23 distance classes presented due to space limitations

<<< Table V Regression (Partial) models>>>

Variable	Model I (Filters)			Model II (linear filters)		
	$\beta$	VIF	t	$\beta$	VIF	t
Predictor Set {A}						
Constant	11.356		427.348***	11.402		443.285***
Size	0.005	1.952	27.172***	0.006	1.592	34.853***
Terrace	-0.325	1.714	-21.534***	-0.294	1.484	-19.75***
Detach	0.208	1.490	11.791***	0.231	1.438	12.629***
Apt	-0.215	1.716	-8.097***	-0.157	1.532	-5.891***
Pre1919	-0.091	2.208	-3.931***	-0.006	1.931	-0.267
Interwar	-0.129	2.622	-6.526***	-0.075	2.258	-3.85***
Post-war	-0.118	2.242	-6.058***	-0.116	2.131	-5.787***
Early-modern	-0.096	1.693	-4.792***	-0.091	1.676	-4.341***
New-Build	0.147	1.142	3.222***	0.114	1.107	2.395**
Reception	0.139	1.633	13.48***	0.134	1.618	13.23***
Predictor Set {B}						
Linear Comb. Filter				0.904	1.147	42.450***
SF n°1	8.547	1.022	29.718***			
SF n°2	-6.488	1.016	-22.632***			
SF n°3	0.174	1.033	0.601			
SF n°4	2.819	1.239	8.903***			
SF n°5	2.997	1.030	10.383***			
SF n°6	-2.614	1.053	-8.953***			
SF n°7	-2.668	1.055	-9.133***			
SF n°8	0.38	1.016	1.324			
SF n°9	-0.871	1.007	-3.052***			
SF n°10	-1.055	1.008	-3.694***			
Total {A}	0.615			0.615		
Total {B}	0.362			0.353		
Total {A+B}	0.767			0.759		
[A.B]{A} only	0.391			0.387		
[A:B] Shared Variance	0.21			0.214		

[B.A]{B} only	0.152	0.139
[1-(A+B)] Unexplain.	0.233	0.241
<i>N</i>	2,664	2,664
<i>R</i> <sup>2</sup>	0.776	0.761
<i>Adj. R</i> <sup>2</sup>	0.769	0.76
<i>F</i>	417.857***	755.25***
<i>AICc</i>	972.408	1291.189

NB. The hold-out model is a Semi-detached, Post 1980 period property. \*\*\*denotes statistical significance at the 1% level; \*\*5% level; \*10% level. NB: Only the first 10 spatial filters are presented due to space limitations. Full spatial filter information is available upon request.

<<<Table VI Interactive (Partial) Regression Model>>>

Variable	Model III (Interactions with Filter)		
	$\beta$	VIF	<i>t</i>
Predictor Set {A}			
Constant	11.259		554.62***
Size	0.005	1.904	28.436***
Terrace*Interwar	-0.325	1.902	-20.6***
Detach*Interwar	0.227	1.536	12.853***
Apt*Interwar	-0.131	1.369	-5.571***
Semi-det*Interwar	0.105	1.003	4.303***
Terrace*Pre1919	0.003	1.485	0.132
Detach*Pre1919	-0.058	1.148	-1.147
Apt*Pre1919	0.403	1.024	1.419
Semi-det*Pre1919	0.018	1.049	0.330
Terrace*Post war	-0.318	1.791	-20.804***
Detach*Post war	0.231	1.672	12.527***
Apt*Post war	-0.137	1.713	-5.209***
Semi-det*Post war	0.105	1.001	4.148***
Terrace*Early modern	-0.037	1.169	-1.327
Detach*Early modern	-0.042	1.206	-1.204
Apt*Early modern	0.029	1.371	0.642
Semi-det*Early modern	0.125	1.002	3.313***
Terrace*New Build	-0.369	1.148	22.762***
Detach*New Build	0.289	1.389	12.97***
Apt*New Build	-0.193	1.019	-2.506**
Semi-det*New Build	0.479	1.001	1.57
Reception	0.167	1.559	14.186***
Predictor Set {B}			
SF n°1	8.736	1.035	30.427***
SF n°2	-6.313	1.039	-21.943***
SF n°3	0.313	1.052	1.082
SF n°4	2.722	1.416	8.105***
SF n°5	3.27	1.099	11.051***
SF n°6	-2.785	1.069	-9.546***
SF n°7	-3.022	1.102	-10.202***
SF n°8	0.274	1.019	0.963

SF n°9	-0.883	1.012	-3.111***
SF n°10	-1.135	1.03	-3.963***
Total {A}	0.634		
Total {B}	0.403		
Total {A+B}	0.781		
[A.B]{A} only	0.377		
[A:B] Shared Variance	0.258		
[B.A]{B} only	0.146		
[1-(A+B)] Unexplain.	0.219		
N	2,664		
$R^2$	0.781		
Adj. $R^2$	0.776		

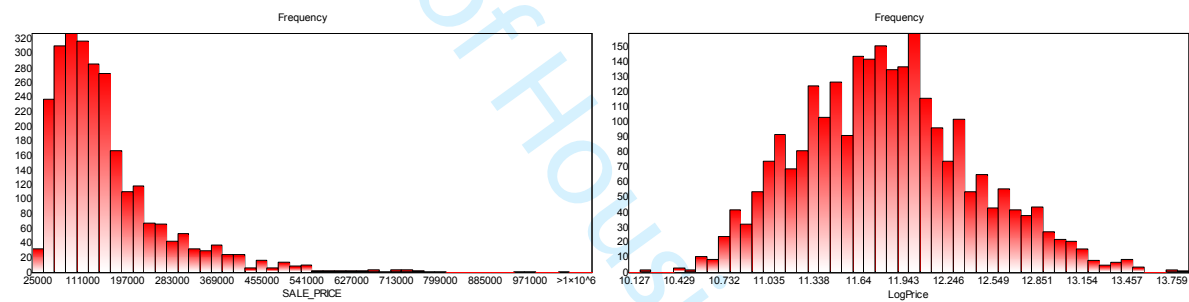
\*\*\*denotes statistical significance at the 1% level; \*\*5% level; \*10% level.

NB: Only the first 10 spatial filters are presented due to space limitations.

Full spatial filter information is available upon request.

## Figures

<<<Figure I Frequency distributions of sale price and the logarithmic of sale price>>>



<<<Figure II Spatial filter diagnostic testing>>>

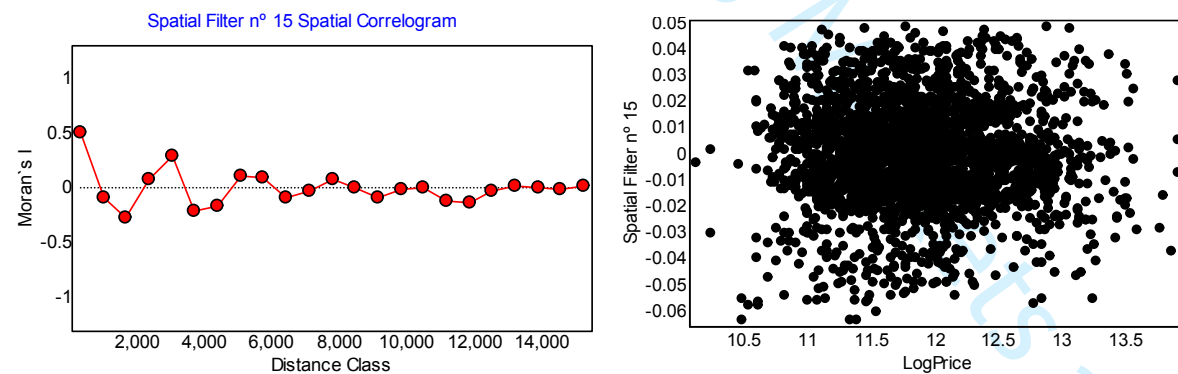
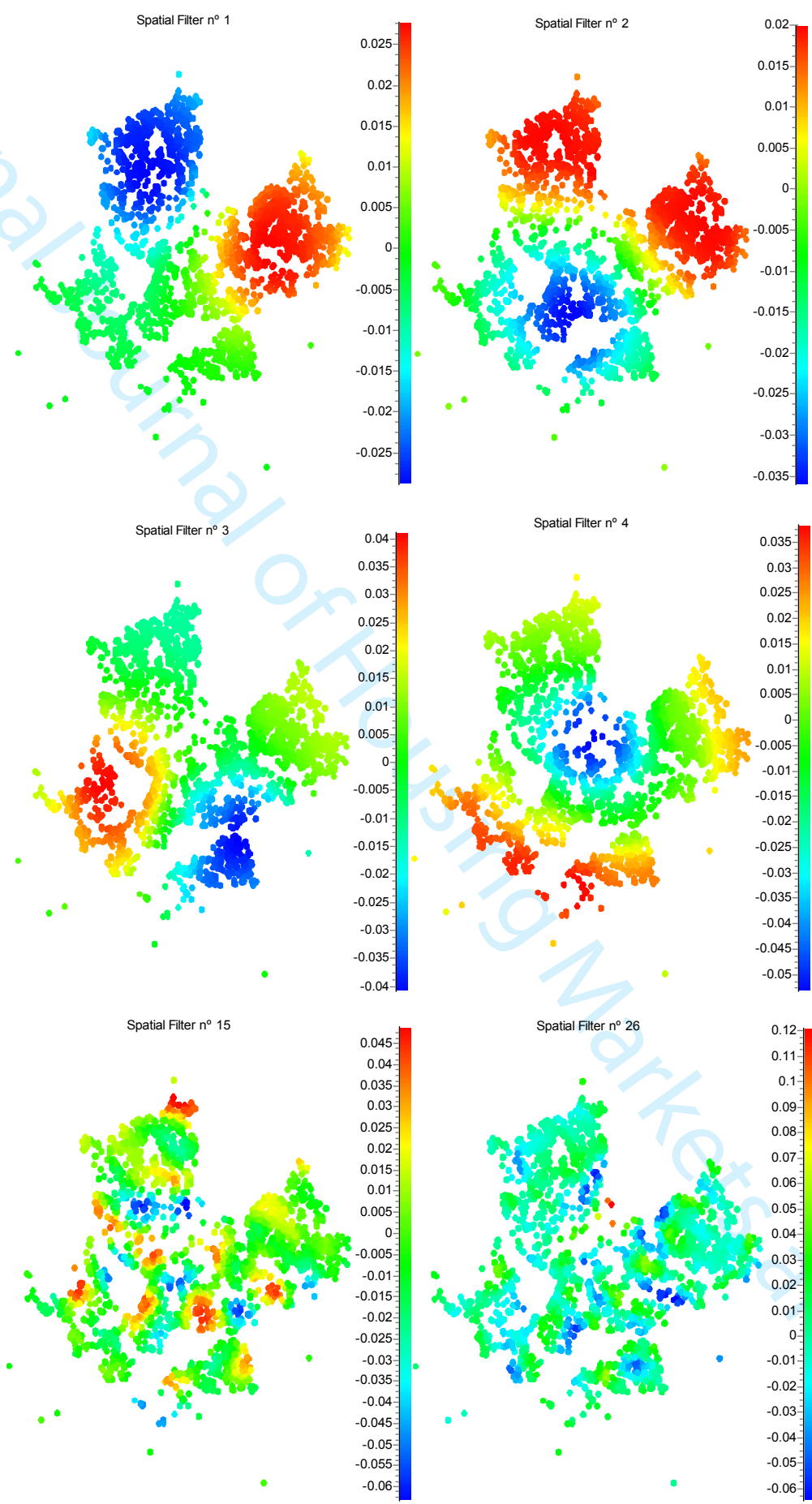
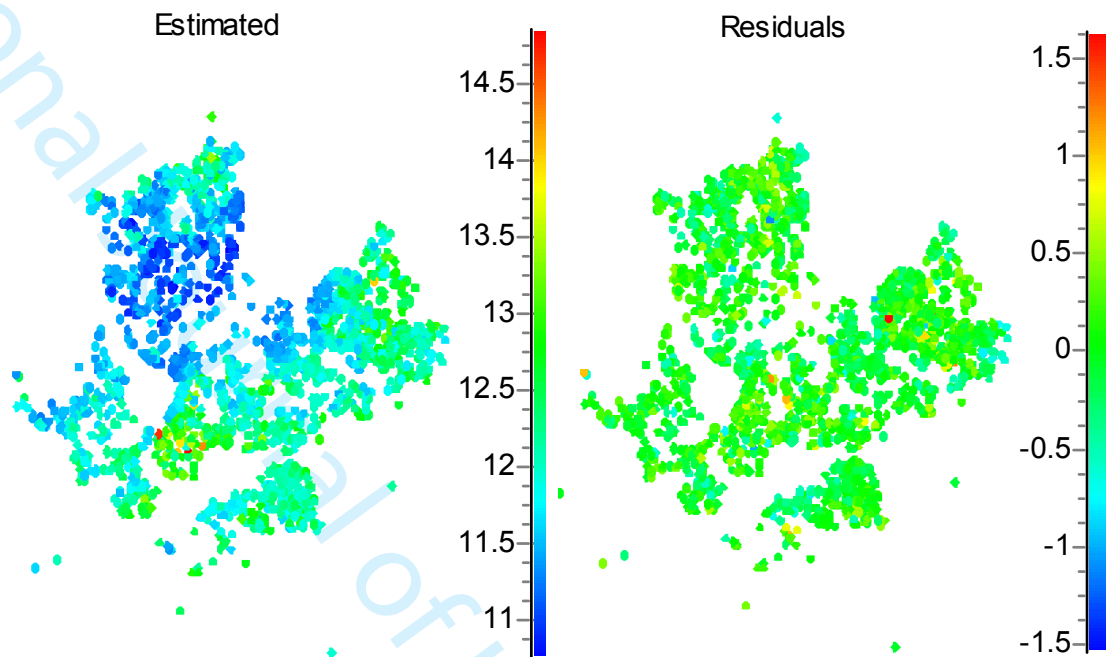
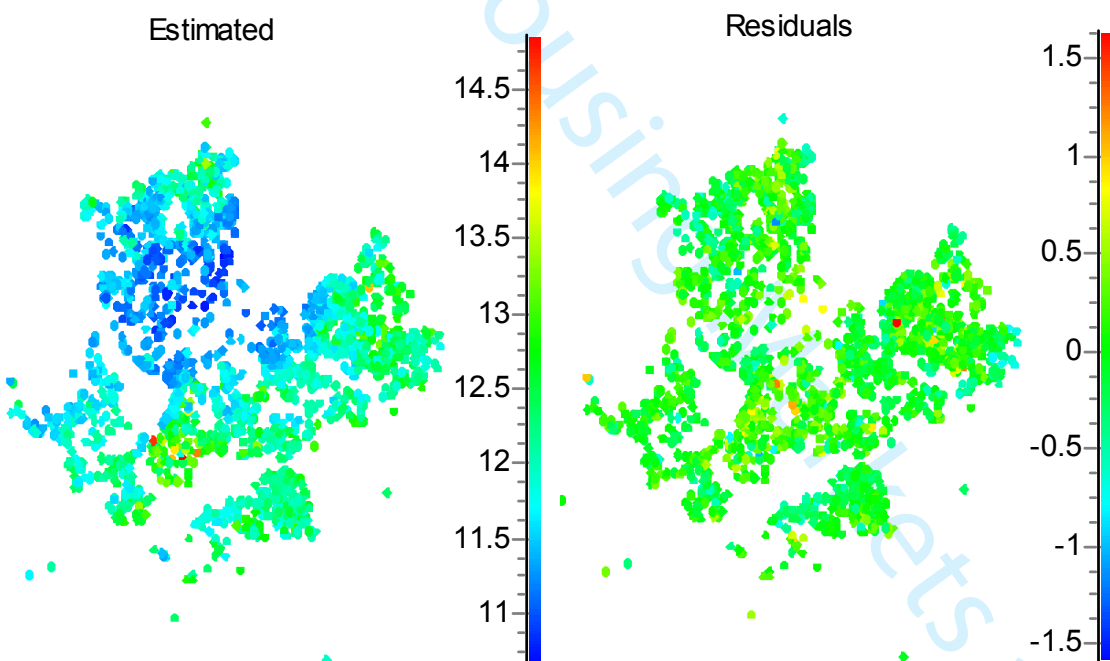


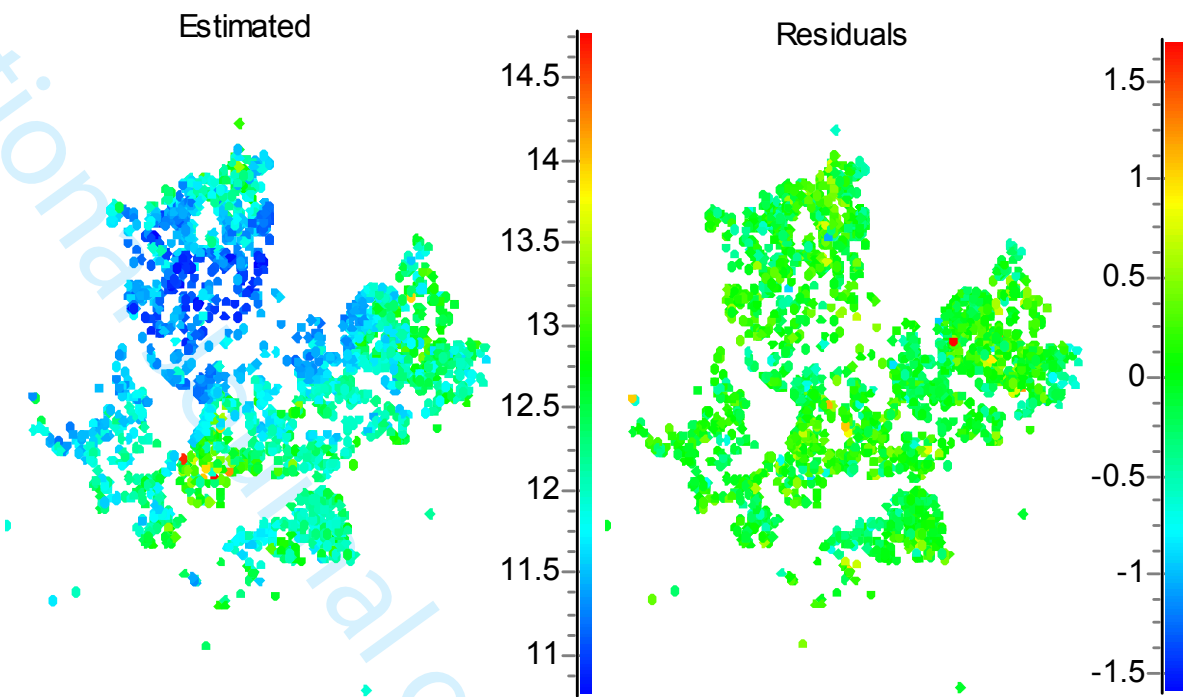
Figure III Spatial Representation of Filters extracted





&lt;&lt;&lt;Figure IV Regression model estimates and residuals&gt;&gt;&gt;

*Model I**Model II**Model III*



## House price estimation using an eigenvector spatial filtering approach

### Abstract

*Purpose:* Numerous geo-statistical methods have been developed to analyse the spatial dimension and composition of house prices. Despite these advances, spatial filtering remains an under-researched approach within house price studies. This paper examines the spatial distribution of house prices using an eigenvector spatial filtering (ESF) procedure, to analyse the local variation and spatial heterogeneity.

*Methods:* Using 2,664 sale transactions over the one year period Q3 2017 to Q3 2018, an eigenvector spatial filtering approach is applied to evaluate spatial patterns within the Belfast housing market. This method consists of using geographical coordinates to specify eigenvectors across geographic distance to determine a set of spatial filters. These convey spatial structures representative of differing spatial scales and units. The filters are incorporated as predictors into regression analyses to alleviate spatial autocorrelation. This approach is intuitive, given that detection of autocorrelation in specific filters and within the regression residuals can be markers for exclusion or inclusion criteria.

*Results:* The findings show both robust and effective estimator consistency and limited spatial dependency - culminating in accurately specified hedonic pricing models. The findings show that the spatial component alone explains 14.6% of the variation in property value, whereas 77.6% of the variation could be attributed to an interaction between the structural characteristics and the local market geography expressed by the filters. This methodological approach reduced short-scale spatial dependency and residual autocorrelation resulting in increased model stability and reduced misspecification error.

*Originality:* Eigenvector-based spatial filtering is a less known but suitable statistical protocol that can be used to analyse house price patterns taking into account spatial autocorrelation at varying (different) spatial scales. This approach arguably provides a more insightful analysis of house prices by removing spatial autocorrelation, both objectively and subjectively, to produce reliable, yet understandable regression models which do not suffer from traditional challenges of serial dependence or spatial mis-specification. This approach offers property researchers and policy makers an intuitive but comprehensible approach for producing accurate price estimation models which can be readily interpreted.

**Keywords:** Eigenvector spatial filtering, Spatially varying coefficients, Hedonic price analysis, house prices, spatial dependency, spatial autocorrelation

### Introduction

There has been increasing emphasis placed on the accuracy of house price estimation and its role in informing urban policy from healthy communities to liveable spaces and connected places. It also retains significance for a wide range of economic activity. In light of this importance, the accuracy, stability and defenceability of house price models is crucial for wider property market analysis and the robust insights into local housing market conditions (Bourassa, Cantoni, and Hoesli, 2010; Seya et al., 2014). Recent advances in geographic information systems (GIS) and geo-statistical and spatial econometric approaches, both parametric and non-parametric, have advanced house price analysis, as they consider the geographically distributed properties of spatial data and identify spatial dependence and spatial

heterogeneity. This has, by-and-large, been a consequence of the awareness of the potential error-bias contained within traditional hedonic models when not acknowledging the spatial heterogeneity of pricing effects (Lu, *et al.*, 2014; Helbich and Griffith, 2015). This increased awareness has propagated considerable interest in accounting for spatial non-stationarity and dependence within hedonic analysis (McCord *et al.*, 2014), challenging the assumption of a constant price function uniformly across a housing market area – an assumption that does not conform to the operation of and complexity of housing market mechanics congruent with urban economic theory (McMillen and Redfearn, 2010).

The advancement of geo-statistical approaches, and the discipline of spatial econometrics, has introduced numerous discussions regarding the confounding effects of space (LeSage and Pace 2009), and primarily the role of spatial autocorrelation – one of the important characteristics of spatial data (Anselin, 1988). This causes the inflation of Type I errors in the significance tests of correlation and regression analyses which can overestimate the degrees of freedom, reduce confidence intervals and result in errors in statistical inference under a null hypothesis, bias coefficients and lead to inappropriate conclusions (LeSage and Pace, 2009; Páez, Fei, and Farber, 2008). Pertinently, many spatial analysis techniques thereby employ model-based statistical inference, the dependability of which is based upon the correctness of assumptions about a model's error term, namely its randomness and independence of observations.

Despite the appealing methodological and practical advantages of these more localised spatial approaches, there remains disagreement within real estate applications as to which method is superlative (Griffith, 2008; Helbich and Griffith, 2015). Numerous approaches have been developed and advocated over the past two decades which have incorporated an eclectic range of spatial effects in order to account for locational effects in hedonic price modelling. These approaches, such as Simultaneous Autoregressive Regression, Spatial Expansion methods, Spatial Lag Models (which incorporate spatial structural instability) and spatial drift models, make use of the spatial characteristics of variables to improve results through reduced error terms and spatial independence (Gao *et al.*, 2006). One prevailing method within property analysis is the Geographically weighted regression (GWR) approach introduced by Fotheringham *et al.*, (1998; 2002), which has become a major approach for explicitly accounting for spatial heterogeneity, using spatially varying coefficients. Nonetheless, this method has not been without its criticisms. Páez *et al.* (2011) and Wheeler and Tiefelsdorf (2005) have indicated that the basic GWR model typically suffers from multicollinearity issues, and also assumes the same degree of spatial smoothness for each coefficient using bandwidth criteria which can affect the model results (Bidanset and Lombard, 2014; Bidanset *et al.*, 2017). There has been refinements to the approach, such as the incorporation of Principal Component Analysis (PCGWR), as a methodology to help remove multicollinearity from inclusionary neighbourhood and locational determinants. More recently there have been augmented approaches suggested incorporating weighting matrices for spatial, temporal and property characteristics, such as the GTCWR approach of Bidanset *et al.* (2018).

One approach which remains relatively unknown, yet emerging as an alternative procedure to address spatial dependence, is the Spatial filtering approach - a method developed in order to obtain enhanced and robust results in a spatial data analysis framework by removing spatial dependency (Griffiths, 1996, 2003; Tiefelsdorf and Griffith, 2007). In its basic format, the eigenvector spatial filtering (ESF) method is an approach that captures spatial dependence applying map pattern variables obtained from spatial connectivity information, using the

Moran (1950) coefficient. This is achieved through the decomposition of a spatial variable/characteristic into trend, a spatially structured stochastic signal, and random noise. In essence, it separates spatially structured random components from both trend and random noise, culminating in leads to sounder statistical inference and useful visualization (Griffith, 1996; 2008; Griffith and Chun, 2014; Helbich and Griffith, 2015). This separation procedure involves eigenfunctions of the matrix version of the numerator of the Moran Coefficient (Griffith and Chun, 2014). Therefore, the application of an ESF approach is to create a spatially structured random component, as captured by a linear combination of selected eigenvectors, in order to mitigate potential error bias through limiting autocorrelation within the residuals.

In this regard, the ESF methodology has been increasingly utilised in a variety of regression settings. The basic model is identical to the standard ordinary (OLS) and generalized (GLS) least squares linear regression models, and therefore, it is easily implemented (Griffith, 2003; Murakami and Griffiths, 2015). As discussed by Tiefelsdorf and Griffith (2007), spatial filtering addresses this from a semi-parametric perspective by generating synthetic explanatory variables reflecting the data's spatial structure. This increases flexibility into model processing in order to analytically decompose a variable into underlying (spatial) components which provides a synthetic variate (the spatial filter) to visualise any spatial autocorrelation contained within a geo-referenced variable (Murakami and Griffiths, 2015). According to Thayn and Simanis (2013) and Franzese and Hays (2014) this approach produces unbiased parameter estimates, reduces spatial misspecification error; increases model fit; increases the normality of model residuals and can increase the homoscedasticity of model residuals spatial dependence and spatial spill-over effect.

This study develops an ESF model (as per Griffith 2003) for the Belfast housing market and thereby models geographically varying relationships using a subset of eigenvectors extracted from a spatial weights matrix as synthetic control variables in a regression model specification. This aims to remove spatial dependence and increase standardised regression models estimation reliability. This approach is seen as furnishing a parsimonious solution to the geographically varying linear regression coefficients problem, a better understanding of multicollinearity, and improved accounting for spatial autocorrelation. More pertinently, it provides professionals with a readily understandable methodology for applying spatial analysis in a more standardised and explainable hedonic framework.

## Literature

Eigenvector Spatial Filtering, as outlined by Murakami and Griffith (2015), has become increasingly popular for the understanding of spatial phenomena, with applications increasing in light of its practicality, readily adaptable process and integration within classical regression based techniques which are considered transparent and understandable. The approach has been adopted across a number of scientific disciplines for understanding spatial interaction, ecological and economic processes and more specifically land-use and housing market analysis. A core stand of this literature has tended to examine the comparative performance of spatial filtering approaches with other locally weighted regression or spatial expansion methodologies and its effectiveness for analysing spatial autocorrelation within regression models.



An early study conducted by Griffith and Peres-Neto (2006), from an ecology perspective, analysed two differing spatial filtering approaches, to help investigate and explain the geographic variability associated with ecological communities. Their results demonstrate the usefulness of eigenfunctions in spatial modelling - specifically that the manifestation of spatial predictors can be easily incorporated into conventional regression models for analysis. Indeed, the authors showed that an important advantage of the spatial filtering methodology over other spatial approaches is that they provide a flexible tool that allows the full range of general and generalized linear modelling theory to be applied to ecological and geographical problems in the presence of nonzero spatial autocorrelation. Analogous findings are evident in the work of Blanchet *et al.* (2008) who, also in an ecological context, investigated the distribution of species, and explicitly the direction of an asymmetric process controlling species distributions along a biogeographical gradient or network, using an eigenfunction-based spatial filtering technique. Comparing the ESF with traditional Moran's eigenvector maps (MEM) analysis within a simulation framework they find that the ESF is superior for producing unbiased coefficient estimations.

With regards to spatial interaction analysis, Chun (2008) tested the assumption of independence among interaction flows engaged in spatial interaction modelling, in the context of U.S. interstate migration flows for measuring network autocorrelation. Undertaking a Stepwise incorporation of eigenvectors, which are extracted from a network link matrix to capture the network autocorrelation in a Poisson regression, the results showed that estimated regression parameters in the spatial filtering interaction model become more intuitively interpretable. Similarly, Fischer and Griffith (2008) compared two approaches, the spatial interaction gravity model and the eigenfunction-based spatial filtering approach, to deal with the issue of spatial autocorrelation amongst flow residuals across 112 European regions. In line with Chun (2008) their findings showed the ESF to be more intuitive. This was also evident in the study of Chun and Griffith (2011) who employed the eigenvector spatial filtering technique to analyse network autocorrelation among migration flows structured through multiple time spans. The findings showed improved model fitting and more intuitive parameter estimates.

In a wider economic context, Crespo-Cuaresma and Feldkircher (2013) also employed spatial filtering to measure the spatial uncertainty of income convergence in Europe using a dataset of income per capita growth and 50 potential determinants for 255 NUTS-2 European regions. The authors reveal that spatial linkages (matrices) comprise an important effect on the estimates of the parameters attached to the model covariates and that income convergence in Europe is influenced by spatially correlated growth spill-overs. Similarly, Patuelli *et al.* (2011) examined regional performance related to unemployment rates in 439 NUTS-3 German districts. They employed a spatial filtering model to unemployment rates in Germany using the derived spatial filters as explanatory variables in a panel modelling framework. Their results show that the computed spatial filters account for most of the residual spatial autocorrelation in the data. In a follow-up study, Patuelli *et al.* (2012) investigate the dynamic adjustment process of unemployment to the study of regional unemployment persistence, in order to account for spatial heterogeneity and/or spatial autocorrelation in both the levels and the dynamics of unemployment. They also employ the use of spatial filtering as a substitute for fixed effects within a panel estimation framework in order to incorporate region-specific information that generates spatial autocorrelation, frees up degrees of freedom and simultaneously corrects for time-stable spatial autocorrelation in the residuals. The authors find widely heterogeneous, but generally high, persistence in regional unemployment rates, signifying that ESF helps provide insights about the spatial patterns in regional adjustment processes.

Griffith and Chun (2014) investigated regional population forecasting for South Korea by incorporating spatial autocorrelation in a generalized linear mixed model framework coupled with eigenvector spatial filtering to capture spatial autocorrelation, namely the complex map pattern portraying spatial dependence that is latent in population counts, and preserves it in regional forecasts of population. The authors find that empirical evaluations of the short run population forecasts indicate that using an ESF to describe spatially structured random effects coupled with a spatially unstructured random effects term furnishes good annual county-level geographic resolution predictions. A further study inspecting regional inequality in China's Guangdong region by Liao and Wei (2015) applied a spatial filtering method in order to eliminate spatial dependence and quantify the extent to which spatial effects have contributed to regional inequality at multiple scales. The results indicated the effect of strengthening spatial dependence with the authors concluding that spatial filtering as a tool helps improve the understanding of complex spatial phenomena.

The approach has also been utilised within the confines of criminal analysis, where Chun (2014) using eigenvector spatial filtering analysed the space-time crime incidents relating to vehicle burglary in Texas, USA, between 2004-2009 within a Poisson generalized linear mixed model specification using ESF. The author shows the approach to be an efficient tool for furnishing robust estimates. In terms of crime mapping and spatial crime analysis, Helbich and Arsanjani (2015) employ ESF as a method for undertaking spatio-temporal mapping to uncover time-invariant crime patterns. Their results suggest that local and regional geography significantly contributes to the explanation of crime patterns. Furthermore, they show annual space-time eigenvectors to indicate spatio-temporal patterns persisting over time. Their findings show that spatial filtering successfully absorbs latent autocorrelation and, therefore, prevents parameter estimation bias whilst increasing the explanatory power of the regression analysis.

Moniruzzaman and Paez (2012) apply spatial filtering for examining urban design analysis and the implications of accessibility to transit for the city of Hamilton, Canada. Employing a logistic regression approach which they highlight are sensitive to overdispersion and spatial error autocorrelation which can result in misleading inference and erroneous policy prescriptions, they show that using spatial filters improved the model inference and accounted for over-dispersion and spatial autocorrelation. In a similar study, Wang, Kockelman and Wang (2013) explored the application of spatial filtering for regression model estimation for transportation land use and land value estimation. Using case studies and appraised values for private properties the authors analysed the effectiveness of spatial filtering in comparison to spatial autoregressive (SAR) models. Their findings showed the SF approach offers increased goodness of fit statistics and more reliable marginal effects of policy variables and other covariates, in comparison to more conventional SAR-based models. Murakami *et al.* (2017) also compare their eigenvector spatially varying coefficient model to GWR to examine land prices for flood hazards in Japan. Using a Monte Carlo simulation technique their study reveals outperformance of geographically weighted regression (GWR) models in terms of the accuracy of parameter estimates and computational time. Further, the authors highlight that the developed model has spatially varying coefficients which have a different degree of spatial smoothness which is a challenge for conventional GWR.

The application of ESF in housing market analysis has been relatively limited, despite that fact that housing policy requires the recognition of spatial heterogeneity in housing prices to account for local settings. McCord *et al.* (2013) examined a number of spatially based modelling frameworks encompassing more traditional approaches (OLS) to more complex

spatial filtering methods to estimate rental values within the Belfast housing market, UK. Their findings revealed that GWR showed increased accuracy, albeit nominal, in predicting marginal price estimates relative to eigenvector filtering and other spatial techniques. Nonetheless, they noted that the high level of segmentation across localised pockets of the housing market needed further analytical insights as the smooth bandwidth did not adequately capture this whereas spatial filtering did, concluding that soft segmentation modelling approaches are essential for understanding rental gradients. More recently, Helbich and Griffith (2016) examined the application of the ESF model in relation to the spatial variation of house prices in a comparative assessment between locally weighted methodologies. The findings showed the ESF to depict a more localized pattern of the parameter estimates without local smoothing. Moreover they revealed the ESF to be less affected by multicollinearity issues between the local parameter estimates than the other approaches. The authors do however show that whilst ESF demonstrates superiority for in-sample explanatory power and prediction accuracy, the weighted regression approaches exhibit slightly better out-of-sample estimations. Nonetheless, they conclude by advocating for the consideration of ESF as a valuable alternative for real estate research that allows going beyond normal probability models.

Overall, the foregoing analysis suggests that spatial filtering comprises relative advantages for understanding complex spatial dependence and autocorrelation across a wealth of ecological and regional economic problems and more latterly housing market analysis. As illustrated by Thayn and Simanis (2013), OLS models whilst comprising well-known limitations for spatial analysis, are useful and easily interpreted, and the assumptions, strengths, and weaknesses of these models are well studied and understood. Accordingly, they advocate that spatial filtering is a powerful geographic method that should be applied to regression based models that use geographic data. In this regard, and in light of a modest number of extant studies examining house prices spatially employing this technique, this paper uses the spatial filtering technique to analyse house price patterns across the Belfast housing market.

**Data and Methodology**

The analysis is conducted using 2,664 sales transactions over the 12 month period (Q3 2017 to Q3 2018) after undergoing a data mining and cleansing exercise to remove outliers. The data was integrated into a GIS platform to append property address information in order to derive absolute location coordinates (*X*, *Y*) required for the spatial modelling exercises<sup>1</sup>. The independent variables are based on the structural characteristics of the properties, including the era of construction, property typology, property size and the number of bedrooms, reception rooms and whether the property comprises a garage (Table I). Where applicable, the categorical variables are transformed into their binary state. This process is undertaken to indicate the absence or presence of a categorical effect that may be expected to shift the outcome (Kleinbaum *et al.*, 1988).

**<<< Insert Table I Variable Descriptions>>>**

A summary of the descriptive statistics for the data is presented in Table II. The sample mean property price is £171,781 which reveals a high dispersion and positive skew (Figure I). As a

<sup>1</sup> The data was exported into SAM, an integrated computational platform tool for spatial analyses (See: Rangel *et al.*, 2010). Processing time for generation of the ESFs equated to approximately 17 minutes for the study sample size. This time accounts for the truncation distance calculation of the maximum connectivity between all sampling units under the minimum spanning tree criterion and the filter selection extraction response determination.

consequence, the logarithmic of sale price was calculated in order to standardise the house price variable and satisfy the statistical assumptions of normality for modelling purposes. The average floor size equates to 106m<sup>2</sup>, again displaying a high variance.

<<<Insert Figure I Frequency distributions of sale price and the logarithmic of sale price>>>

<<<Insert Table II Descriptive Statistics>>>

## Methodology

Initially, topology-based eigenvector based spatial filtering rests upon the seminal work of de Jong, Sprenger, and van Veen (1984), who pioneered studying and applying the relationship between eigenvalues and the Moran's I coefficient to avoid spatial autocorrelation and regression misspecification identified by earlier authors (Cliff and Ord, 1973). In this regard, the ESF method, as developed by Griffith (2000), utilises geographical coordinates which are subject to an eigen analyses of geographical distances to establish a set of spatial filters (eigenvectors) expressing the spatial structure of the region at different spatial scales. In other words, spatial filtering addresses heterogeneity in behaviours through interacting eigenvalues and systematic covariates (Wang *et al.*, 2011). This process exploits eigenvector decomposition techniques, thereby extracting orthogonal and uncorrelated numerical components from a given contiguity matrix (Patuelli *et al.*, 2012), which also emerges in the numerator of the Moran Coefficient statistic. The Moran ESF is based on the Moran coefficient which is a spatial dependence diagnostic statistic formulated as follows:

$$MC = \frac{N \mathbf{y}'\mathbf{M}\mathbf{C}\mathbf{M}\mathbf{y}}{\mathbf{1}'\mathbf{C}\mathbf{1} \mathbf{y}'\mathbf{M}\mathbf{y}} \quad (1)$$

where  $\mathbf{1}$  is an  $N \times 1$  vector of ones,  $\mathbf{y}$  is an  $N \times 1$  vector of variable values,  $\mathbf{C}$  is an  $N \times N$  connectivity matrix whose diagonal elements are zero, and  $\mathbf{M} = \mathbf{I}\mathbf{N} - \mathbf{1}\mathbf{1}'/N$  is an  $N \times N$  matrix for double centring, where  $\mathbf{I}\mathbf{N}$  is an  $N \times N$  identity matrix. Notably,  $\mathbf{M}$  is replaced with  $\mathbf{M}\mathbf{X} = \mathbf{I}\mathbf{N} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  if  $\mathbf{y}$  is a residual vector of a linear regression model. As highlighted by Griffith (2003) and Murakami *et al.* (2017), the MC is positive if the sample values in  $\mathbf{y}$  display positive spatial dependence, and negative if they display negative spatial dependence. The  $l$ -th eigenvector of  $\mathbf{M}\mathbf{C}\mathbf{M}$ ,  $\mathbf{e}_l$ , describes the  $l$ -th map pattern explained by MC, while the set of eigenvectors of  $\mathbf{M}\mathbf{C}\mathbf{M}$ ,  $\mathbf{E}_{full} = \{\mathbf{e}_1, \dots, \mathbf{e}_N\}$ , provides all the possible distinct map pattern descriptions of latent spatial dependence, with each magnitude being indexed by its corresponding eigenvalue (Griffith, 2003).

As illustrated above, the resulting eigenvectors become mutually uncorrelated and orthogonal, with each mimicking a certain degree of latent spatial autocorrelation (SAC), representing global to local patterns (Tiefelsdorf and Griffith, 2007). Accordingly, the eigenvector corresponding to the first eigenvalue,  $\mathbf{e}_1$ , is the set of real values that has the largest positive MC (depicting the maximum positive spatial dependence) achievable by any set of real numbers for the spatial arrangement defined by  $\mathbf{C}$ . The second eigenvector,  $\mathbf{e}_2$ , is the set of real values that has the largest positive MC that is uncorrelated with and orthogonal to  $\mathbf{e}_1$ , and  $\mathbf{e}_N$



is the set of numerical values that has the largest negative MC (depicting the maximum negative spatial dependence) achievable that is uncorrelated with and orthogonal to  $\mathbf{e}_1, \dots, \mathbf{e}_{N-1}$  (Griffith 2003). The set of eigenvectors of  $\mathbf{MCM}$ ,  $\mathbf{E}_{\text{full}} = \{\mathbf{e}_1, \dots, \mathbf{e}_N\}$ , furnishes all possible distinct map pattern descriptions of latent spatial dependence, with each level being indexed by an MC that is proportional to its corresponding eigenvalue. The basic linear model of ESF is:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \quad (2)$$

where  $\mathbf{y}$  is a  $N \times 1$  vector of response variable values;  $\mathbf{E}$  is a  $N \times L$  matrix composed of a subset of  $L$  eigenvectors ( $L < N$ ) from  $\mathbf{E}_{\text{full}}$ ;  $\boldsymbol{\varepsilon}$  is a  $N \times 1$  vector of disturbances;  $\boldsymbol{\beta}$  and  $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_L]$  are parameter vectors whose sizes are  $K \times 1$  and  $L \times 1$ , respectively;  $\sigma^2$  is a variance parameter; and  $\mathbf{0}$  is a  $N \times 1$  vector of zeros. Equation (2) thus is a semiparametric model, with  $\mathbf{M} = \mathbf{I} - \mathbf{1}\mathbf{1}'/N$ , [Eq. (2)] an approximation of a standard spatial lag model, whereas, when  $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ , [Eq. (2)] an approximation of the spatial error model (Tiefelsdorf and Griffith 2007). When  $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ , the eigenvectors are mutually uncorrelated as well as uncorrelated with  $\mathbf{X}$ .

As outlined by Griffith (2003), the  $L$  eigenvectors in  $\mathbf{E}$  are selected by: (1) Eigenvectors representing inconsequential levels of spatial dependence are removed, and (2) significant eigenvectors are chosen using a stepwise selection method. Commonly, step (1) is conducted by removing eigenvectors whose eigenvalues are small or of the wrong nature using the adjusted  $R^2$  as the objective function, with step (2) achieved by maximizing model accuracy or minimizing residual spatial dependence using the MC. Notably, Eq. (2) is identical to the standard linear regression model, thus step (2) can be conducted by using ordinary least squares (OLS) estimation-based stepwise methods. OLS estimators of  $\boldsymbol{\beta}$  and  $\boldsymbol{\gamma}$  are given as:

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{E} \\ \mathbf{E}'\mathbf{X} & \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{E}'\mathbf{y} \end{bmatrix} \quad (3)$$

In addition, Griffith (2008:2761) further augments the basic linear model by introducing interaction terms between the selected eigenvectors and the predictors to model spatially varying coefficients as opposed to using the final EVs to correct for SAC on a global level. Accordingly, the extension takes the following form:

$$\hat{Y} \approx (\beta_0 \mathbf{1} + \sum_{K=1}^{K_0} E_{k0} \beta_{k0}) + \sum_{P=1}^P (\beta_p \mathbf{1} + \sum_{K_p=1}^{K_p} EK_p \beta_{kp}) \cdot X_p + \boldsymbol{\varepsilon} \quad (4)$$

where  $\hat{Y}$  is the  $n \times 1$  vector of prices,  $X_p$  is a  $n \times 1$  vector of independent variable  $p$  ( $p=1,2,3, \dots, P$ ),  $EK_p$  is the  $K_p$  EV ( $k=1,2,3, \dots, K$ ) that describes the variable  $p$ ,  $\beta_0, \beta_{k0}, \beta_{kp}$  are estimated regression coefficients, and  $\boldsymbol{\varepsilon}$  is an independent and identically distributed error term. Note that the element-wise matrix multiplication and the interaction terms are given by  $\beta_{kp} \cdot X_p$ .<sup>2</sup> The first part of the equation represents the spatially varying intercept, and the second part

<sup>2</sup> The parameters are estimated by means of OLS.



represents the spatially varying coefficients. After rearranging, the regression coefficients constitute the global impact, while the individual EVs mimic local modifiers of these global effects across space:

$$Y = \beta_0 1 + \sum_{p=1}^P X_p \cdot 1\beta_p + \sum_{k=1}^K E_k \beta_{E_k} + \sum_{p=1}^P \sum_{k=1}^K X_p \cdot E_k \beta_{pE_k} + \varepsilon \quad (5)$$

In practice, the outlined procedure is challenging due to a large set of covariates and interaction terms, eventually larger than the available number of degrees of freedom. Griffith (2008) originally proposed forward variable selection to find significant interactions, but this procedure is computationally slow (Seya *et al.*, 2014). In order to identify the most relevant interactions in a parsimonious manner, the Akaike information criterion (AIC) is used for evaluation purposes, which considers the model fit and penalizes less parsimonious models. Finally, in order to obtain the final and mappable coefficients, all ESF model parts with common attributes are collected and then factored out in order to determine its spatially varying coefficient (Griffith, 2008).

#### *Partial Regression approach*

The semi-partial regression is used to express the specific portion of variance explained by a given independent variable within the regression analysis (Abdi, 2007). Indeed, this approach is primarily employed for non-orthogonal linear regression to assess the specific effect of each independent variable on the dependent variable (Larsen and McCleary, 1972), where the partial regression coefficient or partial slope coefficient value is dependent upon the other independent variables included in the regression equation. Within the traditional OLS setting, the multiple regression is extended to find a set of partial regression coefficients  $\beta_k$  such that the dependent variable could be approximated as well as possible by a linear combination of the independent variables. Therefore, a predicted value, denoted  $\hat{Y}$ , of the dependent variable is obtained as:

$$\hat{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_k X_k \dots \beta_K X_K \quad (6)$$

The value of the partial coefficients are found using ordinary least squares (OLS). It is often convenient to express the multiple linear regression equation using matrix notation. In this framework, the predicted values of the dependent variable are collected in a vector denoted  $\hat{y}$  and are obtained using:

$$\hat{y} = Xb \text{ with } b = (X^T X)^{-1} X^T y \quad (7)$$

#### **Model development and Eigenvector Filter identification**

As highlighted in the methodology, the ESF approach performs more efficiently in a parsimonious model - as with all spatial modelling architectures. In this regard, there has been continued debate within the field of spatial econometrics relating to the standard statistical testing approach when applied to non-experimental, usually broad-scale spatial data, especially

with respect to model selection procedures (Cohen, 1994). The debate centres around the inclusion of additional parameters for increasing model predictability and the obvious increase in (geo-statistical) model complexity. Consequently, we apply an initial multi-model inference procedure to reduce model complexity, remove potential variance inflation and multicollinearity<sup>3</sup> without compromising model explanation, predictability and stability. This is achieved through the assessment of the minimisation of the Akaike Information Criteria (AICc)<sup>4</sup>. This approach deemed the most parsimonious model, based on 3,095 OLS models tested, to include 12 parameters, excluding the Garage variable and the number of bedrooms thereby providing an  $R^2$  of 63% and Adjusted  $R^2$  of 61.3% (Table III).

<<<Insert Table III OLS Model Selection procedure sorted by Akaike Information Criterion>>>

*Eigenvector Filter identification*

As discerned previously, the purpose of ESF is to remove spatial trends in the response variable, in this instance house prices. The eigenvectors were calculated using geographic coordinates applying a truncated distance function allowing the maximum distance connectivity to connect all sampling units under a minimum tree criterion. The filters were pre-selected based on minimisation of the residual Moran's I ( $p=0.05$ ) across the distance class boundaries (lower and upper) guided by the logarithmic of house price as the response variable. When employed as regressors, the eigenvectors function as proxies for missing explanatory variables, through a parsimonious set of 'candidate' eigenvectors<sup>5</sup>. The residuals obtained constitute the spatially filtered component of the geo-referenced variable examined, computed on the basis of a modified spatial weights matrix. In total, 758 eigenvector filters were established to 'filter' for spatial autocorrelation.

The Eigenvector spatial filtering technique can adopt both a pre-selection criterion and a judicious selection of eigenvectors, as the number of filters appointed tends to increase with both level of linear regression residual spatial autocorrelation and the number of areal units. The spatial filters are subsequently examined with the extraction of the filters to be utilised in the regression modelling undertaken using a filter selection criteria with minimisation of the residuals is achieved based on a local Moran's  $I$  statistic. This automated step for filter pre-selection was further scrutinised to test for potential model customisation regarding the trade-off between increasing the model explanation ( $R^2$ ), the AICs and any potential increases in residual variance inflation. This minimises the residual short-distance spatial autocorrelation and reduces the level of residual autocorrelation. In addition, this step thereby ensured model optimality and model stability whilst further encompassing the assessment of each spatial filters spatial correlogram and the variance of the log-price estimation as demonstrated in Figure II.

<<<Insert Figure II Spatial filter diagnostic testing>>>

<sup>3</sup> This procedure estimates the relative quality of the models for the given set of data, relative to each of the other models premised on the relative information lost by a given model: the less information a model loses, the higher the quality of that model. This therefore estimates the trade-off between the 'goodness of fit' of the model and the simplicity of the model.

<sup>4</sup> The AIC(c) statistic is based on the maximum likelihood of estimating parameters,  $\beta_i$ , where the probability of the observed data would be as large as possible (Burnham and Anderson, 2002), computed as its small sample corrected version as this is asymptotic to the standard version: See De Smith et al. (2007) for a full discussion

<sup>5</sup> selected from the  $n$  eigenvectors, on the basis of their  $MI$  values.

In total, 66 spatial filters were extracted and retained for the regression modelling. As evidenced in Figure III, which displays a sample of extracted filters<sup>6</sup>, each filter extracted presents a detailed representation of the spatial patterns which can have a different degree of spatial structure, smoothness and geographically varying relationship with house prices. Notably the spatial structure becomes more 'localised' when displaying the filters with smaller eigenvalues culminating in more localized parameter surfaces given the reduced truncation distances.

<<<Insert Figure III Spatial Representation of Filters extracted>>>

### OLS and Partial regression results

Having identified the optimal spatial filters using the AIC selection criterion and pragmatic investigative scrutiny of each filter, the spatial filters are used as independent predictors in multiple and partial regression analyses to mitigate spatial autocorrelation and error bias. In total, three regression models are specified to account for location namely, the inclusion of spatial filters (*Model I*) and secondly, the linear combination of the filters suppressed into one location coefficient (*Model II*) and an interaction spatial filter model (*Model III*). Overall, all models display good explanation and performance with an Adjusted  $R^2$  of 0.769, 0.760 and 0.776 with the  $F$ -tests confirming model validity. As observed in Table IV, the distance classifications with their respective accompanying residual Moran's  $I$  value shows that this has reduced to a low level, with only the immediate short-distance (Dis. Class 1) showing the presence of any noteworthy small-scale spatial autocorrelation within the residuals.

<<<Insert Table IV Moran's  $I$  residual autocorrelation within the spatial matrix>>>

Tables V and VI summarise the estimated coefficients across the models. The coefficients for all models infer that for every  $m^2$  increase in property size this equates to a 0.5% and 0.6% increase in price respectively<sup>7</sup>. With regards to property type, both terrace and apartments exhibit statistically significant negative coefficients with the detached coefficient revealing a 23.1% and 25.98% increase ( $p < .01$ ) in models I and II. The property age coefficients reveal negative coefficients across all age categories with the exception of new build properties in both models which exhibits a 14.7% and 11.4% percentage effect. In terms of spatial effects, Model I presents the eigenvector spatial filters which reveal geographically varying regression coefficients, both positive and negative, representing regional patterns and aggregation effects. Model II further presents an overall linear combination of the extracted filters showing the more global coefficient to be significant ( $p < .001$ ).

In terms of partial effects, the OLS model is further constructed to define predictor sets, to examine overlap in explanation under identified predictor set categories. The property characteristics (floor area; type; age) are separated from the spatial filter explanatory parameters to derive a series of additive models which partition the explanation into unique and shared components. For model I, the property characteristics (Predictor set *A*) explain 61.5% of the variation in house prices, with the pure spatial dimension represented by the spatial filters (Predictor set *B*) explaining 36.2% with an overall total explanation of 76.7%. In terms of unique contribution, 39.1% of the variation in property price is explained solely by

<sup>6</sup> the histogram of spatial filter selection residuals evident in Appendix I

<sup>7</sup> Measured using the  $\epsilon(\beta)-1$  transformation as discussed by Halvorsen and Palmquist (1980).

the physical characteristics, with 15.2% exclusively by the spatial characteristics and 21.1% shared explanation between the property characteristics and the spatial component (Table V). For model II, there is a marginally reduced level of explanation for the linear combination of the spatial filters (35.3%) with the unique explanation of the spatial dimension decreasing to 13.9%. However the shared explanation between the predictor sets *A* and *B* marginally increased by 0.4% to 21.4% signalling that the linear combination of the spatial filters nominally increases the level of multicollinearity between the physical and spatial attributes. Overall, the results show that the filter selection inclusion has effectively eliminated residual spatial autocorrelation, whilst not overlapping in any considerable manner with the physical characteristics within the partial regression analysis.

**<<< Insert Table V Regression (Partial) models>>>**

Further refinement of the model specification through the inclusion of interacting property type with age variables, as illustrated in Table VI, slightly increases the model explanatory power, albeit marginally. The findings display an eclectic and varied pricing effect across the property type and age interactions - symbolic of housing market structure, segmentation and heterogeneity, explaining 63.4% of the variation in house prices. In terms of the spatial component, the predictor set B encompassing the spatial filters shows an explanation of 40.3% culminating in an overall  $R^2$  of 78.1%. Further insights as to each specific contribution to the model, 37.7% of the variation in property price is explained solely by the physical characteristics, with 14.6% by the extracted filters with the shared explanation equating to 25.8%.

**<<<Insert Table VI Interactive (Partial) Regression Model>>>**

In terms of spatial representation, Figure IV reveals the estimation and residuals for each respective model. The results present some localised price patterns characteristic of the topographical nature of the housing market structure and two distinctive areas of market segmentation. To the north-west of the Belfast housing market, the results show a lower pricing structure with a number of enclaves towards the centre of this area forming the lowest house prices in the overall market. Towards the south of the market, there is evidence of small pockets of elevated pricing clusters or hot spots. Finally, the model residuals exhibit few instances of elevation and relative stability with limited residual spatial autocorrelation evident across the housing market geography.

**<<<Insert Figure IV Regression model estimates and residuals>>>**

**Conclusion**

Spatial analysis within house price studies has evolved significantly over the past two decades with numerous methodologies having emerged to examine the spatial patterns, heterogeneity and dependency of house prices. This has been fuelled by the ever-growing interest, and indeed importance of housing market policy within urban analyses and policymaking. Despite these advances, eigenvector spatial filtering remains a largely unknown spatial approach within house price estimation studies, despite its increasing application across a variety of other disciplines. The ESF approach provides a foundation for including location within the confines of a traditional regression approach to produce stable, reliable estimates devoid of spatial dependency. This is achieved through the generation of synthetic explanatory variables

reflecting the data's spatial structure. The findings emanating from this study show the effectiveness of applying this methodology to house price sales data for the Belfast housing market, revealing that the spatial filters can be observed as linear combinations of the eigenvectors, and can be regarded as patterns of independent spatial dimensions, culminating in the almost complete elimination of residual spatial autocorrelation and therefore mitigating parameter estimation bias.

The findings exhibit localised parameter surfaces capable of mapping local parameter estimates which does not assume the constant bandwidth or nearest neighbours assumption necessary for other techniques such as GWR, whilst not facing multicollinearity issues between the local coefficients. This provides market professionals and policymakers with a more readily and understandable methodology for applying spatial analysis in a more standardised and explainable hedonic framework. Moreover, the findings, using a partial regression approach in order to isolate the spatial effects, revealed that the unique explanation of the spatial dimension accounts for 14.6% of price variation across the Belfast housing market with limited overlap with the physical characteristics.

The approach has the capacity to aid practitioners in the property taxation field by allowing a decomposition between more intangible spatial elements contributing to value and more tangible physical characteristics. In terms of overall accuracy both are included and defensible. The spatial similarities and differences can be assessed and mapped – depicting similar and dissimilar areas. This can potentially be illustrated by comparable transactions from these areas. The remainder can be illustrated by a more simple model with more understandable parameter estimates for features such as presence of garage or floor area. These estimates will not vary spatially, but across relatively small distances, the cost of providing them also does not vary greatly. This facilitates discussion of the relative contribution of tangible and intangible value factors in a way which may be beneficial for discussions with non-technical consumers of model derived appraisals – such as tax payers and assessment tribunal members. This functionality may be of use as a result in terms of ‘explainability’ of model derived value estimates, which can be crucial in terms of defending assessments and underpinning the operation of effective and efficient property taxes for raising vital public finance.

There is one cautionary note, however. Whilst the ESF approach does offer property researchers and policy makers an intuitive but comprehensible approach for producing accurate price estimation models which can be readily interpreted, one explanation for the ESF approach remaining somewhat of an outcast within house price studies relates to its lack of a ‘user friendly’ interface. As it is premised on eigenvector extraction from a neighbourhood connectivity matrix, this necessitates a large set of interaction terms, as evidenced in this research, which produced 758 eigenvector spatial filters as local parameter estimates. Dealing with this can be both computationally and time intense. It is not, however, outside the bounds of complexity or opacity of comparable spatial and machine learning approaches such as GWR or ANN and perhaps deserves to be included in discussions regarding advanced alternatives to traditional regression analysis for understanding housing markets and for applications seeking to harness such understanding, such as automated valuation modelling for mortgage lending, or mass appraisal of residential values for property taxation purposes.

## References

- Abdi, H. (2007). Part (semi partial) and partial regression coefficients. In. Salkind, N (Ed.) (2007) *Encyclopedia of measurement and statistics*, 736-740, Thousand Oaks (CA): Sage.



- Bidanset, P. E., and Lombard, J. R. (2014). The effect of kernel and bandwidth specification in geographically weighted regression models on the accuracy and uniformity of mass real estate appraisal. *Journal of Property Tax Assessment and Administration*, 11(3), 5-14.
- Bidanset, P. E., Lombard, J. R., Davis, P., McCord, M., and McCluskey, W. J. (2017). Further Evaluating the Impact of Kernel and Bandwidth Specifications of Geographically Weighted Regression on the Equity and Uniformity of Mass Appraisal Models. In *Advances in Automated Valuation Modeling* (pp. 191-199). Springer, Cham.
- Bidanset, P., McCord, M., Lombard, J. A., Davis, P., and McCluskey, W. (2018). Accounting for Locational, Temporal, and Physical Similarity of Residential Sales in Mass Appraisal Modeling: Introducing the Development and Application of Geographically, Temporally, and Characteristically Weighted Regression (GTCWR). *Journal of Property Tax Assessment and*, 14(2), 5-13.
- Blanchet, F. G., Legendre, P., and Borcard, D. (2008). Modelling directional spatial processes in ecological data. *ecological modelling*, 215(4), 325-336.
- Bourassa, S., Cantoni, E., and Hoesli, M. (2010). Predicting house prices with spatial dependence: a comparison of alternative methods. *Journal of Real Estate Research*, 32(2), 139-159.
- Chun, Y. (2008). Modeling network autocorrelation within migration flows by eigenvector spatial filtering. *Journal of Geographical Systems*, 10(4), 317-344.
- Chun, Y., and Griffith, D. A. (2011). Modeling network autocorrelation in space-time migration flow data: an eigenvector spatial filtering approach. *Annals of the Association of American Geographers*, 101(3), 523-536.
- Chun, Y. (2014). Analyzing space-time crime incidents using eigenvector spatial filtering: an application to vehicle burglary. *Geographical Analysis*, 46(2), 165-184.
- Crespo Cuaresma, J., and Feldkircher, M. (2013). Spatial filtering, model uncertainty and the speed of income convergence in Europe. *Journal of Applied Econometrics*, 28(4), 720-741.
- Chun, Y., and Griffith, D. A. (2014). A quality assessment of eigenvector spatial filtering based parameter estimates for the normal probability model. *Spatial Statistics*, 10, 1-11.
- De Jong, P., Sprenger, C., and Van Veen, F. (1984). On extreme values of Moran's I and Geary's c. *Geographical Analysis*, 16(1), 17-24.
- Fischer, M. M., and Griffith, D. A. (2008). Modeling spatial autocorrelation in spatial interaction data: an application to patent citation data in the European Union. *Journal of Regional Science*, 48(5), 969-989.
- Gao, X., Asami, Y., and Chung, C. J. F. (2006). An empirical evaluation of spatial regression models. *Computers and Geosciences*, 32(8), 1040-1051.
- Griffith, D. A. (2000). Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses. *Linear Algebra and its Applications*, 321(1-3), 95-112.
- Griffith, D. A. (1996). Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo-referenced data. *Canadian Geographer/Le Géographe canadien*, 40(4), 351-367.
- Griffith DA (2003) Spatial autocorrelation and spatial filtering: Gaining understanding through theory and scientific visualization. Springer, Berlin.
- Griffith, D. A. (2008). Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment and Planning A*, 40(11), 2751-2769.
- Griffith DA (2009) Modeling spatial autocorrelation in spatial interaction data: empirical evidence from 2002 Germany journey-to-work flows. *J Geogr Syst* 11(2):117-140
- Griffith DA (2010) Spatial filtering. In: Fischer MM, Getis A (eds) Handbook of applied spatial analysis. Springer, Berlin, pp 301-318
- Griffith DA (2011) Visualizing analytical spatial autocorrelation components latent in spatial interaction data: an eigenvector spatial filter approach. *Comput Environ Urban Syst* 35(2):140-149
- Griffith DA, Chun Y (2013) Spatial autocorrelation and spatial filtering. In: Fischer MM, Nijkamp P (eds) Handbook of regional science. Springer, Berlin, pp 1477-1507
- Griffith, D. A. (2009). Spatial autocorrelation. *International encyclopedia of human geography*, 2009, 308-316.

- 1 Griffith, D., and Chun, Y. (2014). Spatial autocorrelation and spatial filtering. *Handbook of regional*
- 2 *science*, 1477-1507.
- 3 Griffith, D. A., and Chun, Y. (2014). An eigenvector spatial filtering contribution to short range regional
- 4 population forecasting. *Economics and Business Letters*, 3(4), 208-217.
- 5 Halvorsen, R. and Palmquist, R. (1980), The interpretation of dummy variables in semi-logarithmic
- 6 equations, *American Economic Review*, Vol. 70 No. 3, pp. 474-475.
- 7 Helbich, M., and Jokar Arsanjani, J. (2015). Spatial eigenvector filtering for spatiotemporal crime
- 8 mapping and spatial crime analysis. *Cartography and Geographic Information Science*, 42(2),
- 9 134-148.
- 10 Helbich, M., and Griffith, D. A. (2016). Spatially varying coefficient models in real estate: Eigenvector
- 11 spatial filtering and alternative approaches. *Computers, Environment and Urban Systems*, 57,
- 12 1-11.
- 13 Jackson, L. E., Kose, M. A., Otrok, C., and Owyang, M. T. (2016). Specification and estimation of
- 14 bayesian dynamic factor models: A monte Carlo analysis with an application to global house
- 15 price comovement. In *Dynamic Factor Models* (pp. 361-400). Emerald Group Publishing
- 16 Limited.
- 17 Larsen, W. A., & McCleary, S. J. (1972). The use of partial residual plots in regression analysis.
- 18 *Technometrics*, 14(3), 781-790.
- 19 LeSage J.P. and Pace R. K. (2009) Introduction to spatial econometrics. Chapman and Hall/CRC, Boca
- 20 Raton
- 21 LeSage J.P. and Pace R. K. (2014) The biggest myth in spatial econometrics. *Econom* 2(4):217-249.
- 22 Liao, F. H., and Wei, Y. D. (2015). Space, scale, and regional inequality in provincial China: A spatial
- 23 filtering approach. *Applied Geography*, 61, 94-104.
- 24 Maza, A., and Villaverde, J. (2008). A note on the need to account for spatial dependence: A case of
- 25 migratory flows in Spain. *The review of regional studies*, 38(1), 105.
- 26 McCord, M., Davis, P. T., Haran, M., McIlhatton, D., and McCord, J. (2014). Understanding rental
- 27 prices in the UK: a comparative application of spatial modelling approaches. *International*
- 28 *Journal of Housing Markets and Analysis*, 7(1), 98-128.
- 29 Murakami, D., Yoshida, T., Seya, H., Griffith, D. A., and Yamagata, Y. (2017). A Moran coefficient-
- 30 based mixed effects approach to investigate spatially varying relationships. *Spatial Statistics*,
- 31 19, 68-89.
- 32 Moniruzzaman, M., and Páez, A. (2012). Accessibility to transit, by transit, and mode share: application
- 33 of a logistic model with spatial filters. *Journal of Transport Geography*, 24, 198-205.
- 34 Páez, A., Farber, S., and Wheeler, D. (2011). A simulation-based study of geographically weighted
- 35 regression as a method for investigating spatially varying relationships. *Environment and*
- 36 *Planning A*, 43(12), 2992-3010.
- 37 Páez, A., Long, F., and Farber, S. (2008). Moving window approaches for hedonic price estimation: an
- 38 empirical comparison of modelling techniques. *Urban Studies*, 45(8), 1565-1581.
- 39 Patuelli, R., Schanne, N., Griffith, D. A., and Nijkamp, P. (2012). Persistence of regional
- 40 unemployment: Application of a spatial filtering approach to local labor markets in Germany.
- 41 *Journal of Regional Science*, 52(2), 300-323.
- 42 Patuelli, R., Griffith, D. A., Tiefelsdorf, M., and Nijkamp, P. (2011). Spatial filtering and eigenvector
- 43 stability: space-time models for German unemployment data. *International Regional Science*
- 44 *Review*, 34(2), 253-280.
- 45 Rangel, T.F.; Diniz-Filho, J.A.F. and Bini, L.M. (2010), SAM: A comprehensive application for Spatial
- 46 Analysis in Macroecology. *Ecography*, 33: 1-5.
- 47 Seya H, Tsutsumi M, Yamagata Y (2014a) Weighted-average least squares applied to spatial
- 48 econometric models: a Monte Carlo investigation. *Geographical Analysis* 46(2):126-147
- 49 Seya, H., Tsutsumi, M., Mendez, S. A., and Vega, A. M. (2012). Application of model averaging
- 50 techniques to spatial hedonic land price models. *Econometrics: New Research*, 63-88.
- 51 Seya, H., Murakami, D., Tsutsumi, M., and Yamagata, Y. (2015). Application of LASSO to the
- 52 Eigenvector Selection Problem in Eigenvector-based Spatial Filtering. *Geographical Analysis*,
- 53 47(3), 284-299.
- 54
- 55
- 56
- 57
- 58
- 59
- 60

Thayn, J. B., and Simanis, J. M. (2013). Accounting for spatial autocorrelation in linear regression models using spatial filtering with eigenvectors. *Annals of the Association of American Geographers*, 103(1), 47-66.

Tiefelsdorf, M., and Griffith, D. A. (2007). Semiparametric filtering of spatial autocorrelation: the eigenvector approach. *Environment and Planning A*, 39(5), 1193-1221.

Wang, Y., Kockelman, K. M., and Wang, X. C. (2013). Understanding spatial filtering for analysis of land use-transport data. *Journal of Transport Geography*, 31, 123-131.

Wheeler, D., and Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems*, 7(2), 161-187.

Appendix I

Histogram of spatial filter selection residuals

